



Facultad de Ciencias Bioquímicas y Farmacéuticas

Universidad Nacional de Rosario

*Tesis para optar a título de Doctor en Ciencias Biológicas*

# “Interacción de las proteínas de procesamiento con los precursores de miARN y su mecanismo en plantas”

*Irina Paula Suarez*

Rosario, Argentina

2015

# “Interacción de las proteínas de procesamiento con los precursores de miARN y su mecanismo en plantas”

Irina Paula Suarez

Licenciada en Biotecnología

Universidad Nacional de Rosario

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Químicas de la Universidad Nacional de Rosario y no ha sido presentada previamente para la obtención de otro título en esta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET-UNR) y en el Área Biofísica del Departamento de Química Biológica, dependiente de la Facultad de Ciencias Bioquímicas y Farmacéuticas, durante el período comprendido entre el 1ro de abril de 2011 y el 30 de noviembre de 2015, bajo la dirección del Dr. Rodolfo M. Rasia.

*A mi abuelo Armando*

## Agradecimientos

Agradezco profundamente haber tenido la oportunidad de transitar esta experiencia. Aún más agradezco haber compartido todo esto con tantas personas maravillosas. Principalmente agradezco a mi director, Fito Rasia, por su inmensa generosidad y paciencia.

A mis muy geniales compañeros de grupo Pau, Fer M., Guille H., Flor M., Robi, Nah, Fer M. II, Diegui, Salva. Y anexos Markus, Maru, Guille B., Lis, Andru, Anto, Marian G., Bruni, Moyi, Flor E., Ari, Sol, Luchi, Pabli, Carola, Carli, Marian, Ane, Pau H. A todos por su alegría, amistad y compañía.

A mi familia por su apoyo incondicional y la libertad. Especialmente a mi abuela Chiche por su sabiduría y los recuerdos, y a mi abuelo Armando, a quien debo y dedico mi amor por la vida todos los días.

A mis amados amigos, La Banda y Los Chanchis, por el andar juntos. Especialmente Agus, Dani, Cari, Mati y Pau por su presencia en los últimos tiempos.

Agradezco a los Doctores Claudia Spampinato y Luis M. Veggi por su trabajo como tutores.

A los Doctores Alejandro Tato Vila, Javier Palatnik y Jèrôme Boisbouvier por su colaboración con nuestro grupo.

A mi directora alemana, Dra. Claudia Höbartner, por abrirme las puertas de su laboratorio en el Instituto Max Planck de Göttingen.

A CONICET y la Agencia Nacional de Promoción Científica y Tecnológica por el apoyo económico.

Al Servicio Alemán de Intercambio Académico (DAAD) y la Fundación Bunge y Born de Argentina por las becas otorgadas para mi trabajo en Alemania.

# Contenido

Contenido .....	i
Publicaciones .....	1
Abreviaturas .....	2
Resumen .....	4
Abstract .....	5
CAPITULO 1: Introducción al procesamiento de miARNs .....	6
1.1. Silenciamiento génico .....	6
1.1.1. Descubrimiento del ARN de interferencia .....	6
1.1.2. Silenciamiento génico post-transcripcional por ARNs pequeños en plantas .....	7
1.2. Biogénesis de ARNs pequeños en plantas .....	8
1.2.1. siARNs, tasiARNs y nat-siARNs .....	8
1.2.2. micro-ARNs .....	10
1.3. Biogénesis de miARNs en animales vs. plantas .....	12
1.4. Determinantes estructurales en precursores de miARN .....	13
1.5. Enzimas ARNasa III, bases estructurales del reconocimiento del sustrato .....	15
1.5.1. Clase I: ARNasa III .....	16
1.5.2. Clase II: Drosha .....	18
1.5.3. Clase III: Dicer .....	19
1.5.4. Enzimas tipo Dicer en plantas .....	21
1.6. Relevancia de los dsRBDs en la biogénesis de miARNs .....	22
1.7. Reconocimiento del sitio de corte sobre precursores de miARNs .....	25
1.8. Objetivos .....	26
2. CAPITULO 2: Materiales y Métodos .....	27
2.1. Cepas, plásmidos y medios de cultivo .....	27
2.1.1. Cepas bacterianas .....	27
2.1.2. Genes sintéticos .....	27
2.1.3. Vectores plasmídicos .....	27
2.1.4. Medios de cultivo .....	27
2.2. Transformación bacteriana .....	28
2.2.1. Producción de células competentes .....	28
2.2.2. Transformación de <i>E. coli</i> .....	28
2.2.3. Selección y análisis de transformantes .....	28
2.3. Subclonado de fragmentos de ADN .....	28
2.3.1. Minipreparaciones de ADN plasmídico .....	28

2.3.2.	Estimación de la concentración de ADN .....	28
2.3.3.	Digestión de ADN con endonucleasas de restricción .....	29
2.3.4.	Electroforesis de ADN en geles de agarosa.....	29
2.3.5.	Purificación de fragmentos de ADN a partir de geles de agarosa .....	29
2.3.6.	Ligación de fragmentos de ADN.....	29
2.4.	Expresión y purificación de proteínas.....	29
2.4.1.	Expresión de proteínas recombinantes .....	29
2.4.2.	Replegado de proteínas .....	31
2.4.3.	Purificación de la proteasa TEV.....	31
2.4.4.	Estimación de la concentración de proteínas .....	32
2.4.5.	Electroforesis de proteínas en geles de poli(acrilamida).....	32
2.5.	Espectroscopia de resonancia magnética nuclear .....	32
2.5.1.	Fundamentos teóricos .....	32
2.5.2.	Procedimientos experimentales .....	37
2.6.	Preparación de muestras de ARN .....	39
2.6.1.	Purificación de oligonucleótidos de ADN .....	39
2.6.2.	Precipitación de ácidos nucleicos con etanol.....	39
2.6.3.	Moldes para transcripción <i>in vitro</i> .....	39
2.6.4.	Extracción Fenol / Cloroformo / Isoamílico.....	41
2.6.5.	Transcripción <i>in vitro</i> a escala intermedia.....	41
2.6.6.	Transcripción <i>in vitro</i> a pequeña escala .....	41
2.6.7.	Determinación de la concentración de muestras de ácidos nucleicos .....	41
2.7.	Ensayos de mapeo estructural sobre ARN.....	42
2.7.1.	Marcado de muestras de ARN con <sup>32</sup> P en extremo 5' .....	42
2.7.2.	Mapeo por digestión enzimática .....	42
2.7.3.	Mapeo por hidrólisis inducida por Terbio (3+).....	42
2.7.4.	Geles de Secuenciación.....	43
2.8.	Titulaciones por anisotropía de la fluorescencia .....	43
2.9.	Dicroísmo circular .....	44
2.10.	Programas y servidores utilizados .....	44
3.	CAPITULO 3: DCL1-A, primera aproximación.....	47
3.1.	DCL1-A es una proteína desestructurada .....	48
3.2.	DsRBDs similares a DCL1-A .....	49
3.2.1.	DsRBDs de estructura conocida .....	49
3.2.2.	DsRBDs de las proteínas tipo Dicer de <i>A. thaliana</i> .....	51
3.3.	Afinidad por el sustrato .....	53

3.4.	DCL1-A se pliega al interactuar con el sustrato .....	56
3.4.1.	Plegamiento por unión a ARNdh.....	56
3.4.2.	La unión de DCL1-A a ADN no induce su plegamiento.....	59
3.5.	Discusión.....	61
4.	CAPITULO 4: ¿Cuán desplegada se encuentra DCL1-A libre? .....	64
4.1.	DCL1-A no está completamente desplegada .....	65
4.2.	Grado de compactación.....	65
4.3.	DCL1-A explora conformaciones de hélice $\alpha$ en el C-terminal.....	67
4.3.1.	Asignación del esqueleto carbonado .....	67
4.3.2.	Estructura secundaria parcial a 298 K.....	68
4.4.	Dinámica de la cadena de DCL1-A .....	70
4.5.	Estabilización de la estructura secundaria a 278 K .....	71
4.6.	Orientaciones en el espacio preferidas por DCL1-A .....	72
4.7.	DCL1-A libre: un ensamble de múltiples conformaciones en intercambio .....	74
4.7.1.	Construcción del ensamble conformacional de DCL1-A .....	74
4.7.2.	Determinación de subensambles de conformeros seleccionados por un algoritmo genético .....	78
4.7.3.	Validación cruzada de la calidad del ensamble conformacional .....	79
4.7.4.	Descripción estructural del ensamble conformacional de DCL1-A .....	80
4.8.	Discusión.....	83
5.	CAPITULO 5: DCL1-A en complejo con el sustrato, mecanismo de unión y plegamiento....	85
5.1.	DCL1-A adquiere un plegamiento tipo dsRBD en presencia del sustrato .....	85
5.1.1.	Asignación del esqueleto peptídico de DCL1-A en complejo con ARNdh .....	85
5.1.2.	Determinación de la estructura secundaria adquirida.....	86
5.1.3.	Cálculo de la estructura de DCL1-A en complejo con ARNdh .....	87
5.2.	Mecanismo de unión al sustrato y plegamiento.....	91
5.2.1.	DCL1-A plegada se encuentra en intercambio con la forma desplegada.....	91
5.2.2.	La forma desplegada de DCL1-A no desaparece en presencia de exceso de sustrato .....	92
5.2.3.	El mecanismo de unión al sustrato y plegamiento involucra una especie intermediaria .....	93
5.3.	El intermediario del mecanismo de unión explora conformaciones intermedias.....	96
5.4.	Discusión.....	99
6.	CAPITULO 6: Arquitectura de los precursores de micro-ARNs.....	101
6.1.	Mapeo estructural por digestión enzimática .....	103
6.2.	Mapeo de flexibilidad por hidrólisis inducida por Terbio (III) .....	113
6.3.	Discusión.....	115

7. Conclusiones finales .....	116
8. APENDICE.....	117
8.1. Apéndice I.....	117
8.2. Apéndice II.....	123
Bibliografía.....	128



## Publicaciones

Parte de los trabajos realizados en esta tesis han sido presentados en las siguientes publicaciones:

Burdisso P, Suarez IP, Bologna NG, Palatnik JF, Bersch B, y Rasia RM (2012) "Second dsRBD of DCL1: Structure and nucleic acid binding", *Biochemistry*, (51): 10159–10166.

Suarez IP, Burdisso P, Benoit MPMH, Boisbouvier J, y Rasia RM (2015) "Induced folding in RNA recognition by *Arabidopsis thaliana* DCL1", *Nucl. Acids Res.*, 43 (13): 6607-6619.

## Abreviaturas

ADN Ácido desoxirribonucleico.

ADNasa ADN nucleasa.

ADNc ADN copia.

AGO Proteína de la familia ARGONAUTA.

ARN Ácido ribonucleico.

ARNasa ARN nucleasa.

ARNdh ARN doble hebra.

ARNi ARN de interferencia.

ARNm ARN mensajero.

ARNsh ARN simple hebra.

CBC Complejo de unión a caperuza.

CD Dicroísmo Circular.

DCL Ribonucleasa III tipo Dicer.

DGCR8 Di George syndrome critical region gene 8.

dsRBD Dominio de unión a ARNdh.

EXP5 Exportin-5.

HEN1 Hua enhancer 1.

HSQC Experimento de coherencia cuántica simple heteronuclear.

HYL1 Hyponastic leaves 1.

IDP Proteína Intrínsecamente Desordenada.

LB Luria-Bertani.

M9 Medio mínimo M9.

MiARN Micro-ARN.

Min Minuto/s.

nat-siARN siARNs naturales.

Nt Nucleótidos.

Pb Pares de base.

PCR Reacción en cadena de la polimerasa.

PKR Proteína Kinasa dependiente de ARN.

Pol II ARN polimerasa II.

RDC Acoplamientos Dipolares Residuales.

RDR ARN polimerasas dependientes de ARN.

Rh Radio Hidrodinámico.

RISC Complejo de silenciamiento inducido por ARN.

RMN Resonancia Magnética Nuclear.

RMSD Raiz Cuadrada del Desvío Medio.

RT Retrotranscripción.

SE Serrate.

seg Segundo/s.

siARN ARN pequeño de interferencia.

tasiARN siARN de accion en trans.

## Resumen

DCL1 es la enzima ribonucleasa central en la biogénesis de micro-ARNs en plantas. Su función consiste en realizar dos cortes sucesivos sobre transcritos primarios para liberar el micro-ARN maduro. El mecanismo a través del cual la enzima reconoce los sitios de corte sobre los precursores resta ser elucidado. La enzima presenta dos dominios de unión a ARN doble hebra en tándem en el extremo carboxilo-terminal, los cuales son esenciales para su funcionamiento *in vivo*.

En este trabajo de tesis nos dedicamos principalmente a la descripción biofísica del primero de los dominios de unión a ARN doble hebra de DCL1 de *Arabidopsis thaliana*, el cual se encuentra altamente conservado entre las distintas especies de plantas. Demostramos que el dominio libre en solución se encuentra desestructurado tanto en forma aislada como en presencia de regiones vecinas de la proteína. A pesar de estar desestructurado, es capaz de unir ARN sustrato. Mediante el análisis estructural del complejo con el ARN encontramos que el dominio adquiere una conformación plegada al entrar en contacto con el sustrato. La topología de la conformación plegada adquirida, que calculamos en base a datos de RMN, se corresponde con la esperada para un dominio de la familia de los dominios de unión a ARN doble hebra. Una vez definidas las características de este sistema procedimos a realizar, por un lado, una descripción más profunda de la forma libre desestructurada y, por otro, proponer un mecanismo posible para la unión al sustrato y plegamiento de la proteína.

De la forma libre podemos decir que globalmente es una especie flexible, aunque no se encuentra completamente desestructurada. A través de distintos observables de RMN logramos determinar que la región correspondiente al extremo carboxilo-terminal presenta una cierta rigidez y explora conformaciones tipo hélice alfa. La presencia de esta estructura preformada podría resultar esencial para la unión al sustrato y el plegamiento.

En cuanto al mecanismo de unión y plegamiento pudimos obtener evidencia que sugiere la existencia de un complejo intermediario entre la forma libre desplegada y la forma plegada unida al ARN. En este complejo intermediario la proteína se encuentra en una tercer especie, también mayormente desplegada y unida al sustrato.

Finalmente en una tercera etapa del trabajo nos abocamos al estudio desde el punto de vista estructural de precursores completos de micro-ARNs. Buscamos validar los determinantes que permiten reconocer la posición del micro-ARN dentro del precursor por parte del complejo de procesamiento formado por DCL1 y las proteínas accesorias HYL1 y SERRATE. Estudiamos cuatro precursores de diferentes familias por técnicas bioquímicas de sondeo de estructura secundaria y flexibilidad de la cadena. Logramos validar experimentalmente determinantes estructurales propuestos a partir de estructuras calculadas *in silico*. También encontramos indicios que sugieren que la posición del sitio del primer corte, el cual es fundamental para definir el registro de los cortes sucesivos ejecutados por DCL1, podría estar relacionada con un cambio abrupto desde una región de la hebra con una conformación flexible a una de menor flexibilidad. Esta particularidad podría estar permitiendo una interacción específica con las proteínas que conforman el complejo de procesamiento.

## Abstract

DCL1 is the main ribonuclease in micro-RNA biogenesis in plants. Its function consists in performing two successive cuts on primary transcripts in order to release the mature micro-RNA. The mechanism through which the enzyme recognizes the cleavage sites remains to be elucidated. The enzyme presents two double stranded RNA binding domains in tandem on the carboxylic-terminal end, which are essential for its *in vivo* function.

In this work, we focused mainly on the biophysical description of the first double stranded RNA binding domain of DCL1 of *Arabidopsis thaliana*, which is highly conserved among plant species. We have demonstrated that the free domain in solution is unstructured, not only when isolated but also when surrounding regions of the protein are present. Even though it is unstructured, the domain is capable of binding to the substrate. Through the structural analysis of the complex with the RNA we found the domain acquires a folded conformation when it binds the substrate. The topology of the acquired folded conformation, calculated based on NMR data, corresponds to the double stranded RNA binding domain family. Once we defined these characteristics on the system we proceeded to perform a deeper characterization on the free unstructured form and to propose a possible mechanism for substrate binding and folding of the protein.

On the free form we found that it is globally a flexible species, even though it is not completely unstructured. Through different NMR observables we have determined that the carboxylic-terminal end presents a certain rigidity and explores alfa helix conformations. This prefolded structure could be essential for binding to the substrate and folding.

On the binding and folding mechanism we could obtain evidence that suggests the existence of an intermediate complex between the free unfolded form and the RNA bound folded form. In this intermediate complex the protein is also mainly unfolded but bound to the substrate.

Finally, on a third stage of this work we performed a structural study of full-length micro-RNA precursors. We have focused on validating the determinants that allow the recognition of the micro-RNA position in the precursor by the processing complex involving DCL1 and the accessory proteins HYL1 and SERRATE. Employing biochemical techniques for probing secondary structure and strand flexibility we have studied four precursors from different families. We were able to experimentally validate structural determinants proposed on the basis of *in silico* calculated structures. We found evidence that suggests the position of the first cleavage site, which is essential for defining the register for the successive DCL1 cuts, could be related with an abrupt change from a flexible region of the strand to a much lesser flexible one. This peculiarity could be allowing for a specific interaction with the proteins involved in the processing complex.

# 1. CAPITULO 1: Introducción al procesamiento de miARNs

## 1.1. Silenciamiento génico

### 1.1.1. Descubrimiento del ARN de interferencia

El ARN de interferencia (ARNi) es utilizado por las células para silenciar la actividad de genes específicos. Los primeros pasos en el descubrimiento del fenómeno del ARNi fueron dados en 1990 por Napoli y Jorgensen (Napoli et al. 1990), mientras estudiaban la vía de síntesis de pigmentos de flores de petunia. Con el objetivo de obtener una coloración violeta más intensa, estos investigadores sobreexpresaron en petunias silvestres el gen de la chalcona sintasa (CHS), una enzima clave en la síntesis de flavonoides. Sorpresivamente, hallaron que algunas de las flores obtenidas eran blancas y que los niveles de CHS en éstas eran 50 veces menores que en petunias salvajes. En 1992, Romano y Marcino reportaron un fenómeno similar en *Neurospora crassa* (Romano & Macino 1992). Notaron que la introducción de secuencias de ARN homólogas a un gen endógeno, causaban la represión de su expresión. Los primeros reportes de silenciamiento génico en animales se obtuvieron en el año 1995 en un trabajo realizado en *Caenorhabditis elegans* (Guo & Kemphues 1995). En el mismo, Guo y Kemphues observaron que la introducción de ARN antisentido al ARN mensajero (ARNm) del gen *par-1* resultaba en la degradación del transcripto correspondiente. De manera inesperada, cuando se realizaron experimentos control usando sólo el ARN sentido de *par-1*, que es incapaz de hibridar con el transcripto de *par-1* endógeno, observaron que el mensajero de este gen seguía siendo blanco de la degradación. En 1998, Fire y Mello publicaron un trabajo en el que demostraron que la inyección de ARN doble hebra (ARNdh), una mezcla de ARN sentido y ARN antisentido, en *C. elegans* era mucho más eficiente en el silenciamiento del gen que la inyección de ARN simple hebra (ARNsh) (Fire et al. 1998). En el mismo trabajo, también demostraron que la inyección de unas pocas moléculas de ARNdh por célula era suficiente para silenciar la expresión del gen. Más interesante aún, fue el descubrimiento de que la inyección de ARNdh en el intestino del gusano provocaba que el silenciamiento se dispare en todo el organismo e, incluso, se transmita a la primera generación de descendientes. Este trabajo, publicado en la revista *Nature*, dio el puntapié inicial en el estudio y comprensión de la regulación de la expresión génica mediante ARNs pequeños. En el año 2006 los autores del trabajo fueron reconocidos por la Academia Real Sueca de Ciencias al recibir el Premio Nobel de Fisiología y Medicina. Al hallar que la inyección de ARNdh en un gusano o planta causaba un silenciamiento sistémico y que, además, era transmitido a la siguiente generación los investigadores sospecharon que este mecanismo estaba mediado por alguna molécula efectora del silenciamiento. Los primeros estudios en la búsqueda de estos intermediarios fueron realizados en 1999 por Hamilton y Baulcombe, mientras estudiaban el fenómeno de silenciamiento génico post-transcripcional inducido por virus en plantas (Hamilton 1999). Si bien en ese momento existía la idea general de que el ARNdh tenía que separarse para que la hebra antisentido se una al ARNm, la hebra antisentido de longitud completa nunca fue detectada. Este hecho, llevó a Hamilton y Baulcombe a buscar formas más cortas del ARN antisentido derivado del ARNdh. La hipótesis de estos investigadores consistía en que el ARN antisentido podría servir como una guía, uniéndose al ARNm complementario y causando su degradación. Al descubrir que el ARN antisentido tenía una longitud estimada de 25 nucleótidos (nt), sugirieron que esta cantidad de nt era la necesaria para la especificidad del ARNi. Al año siguiente, dos grupos independientes identificaron ARNs de 21-23 nucleótidos que copurificaban con el ARNi, sugiriendo que el ARNdh era convertido en

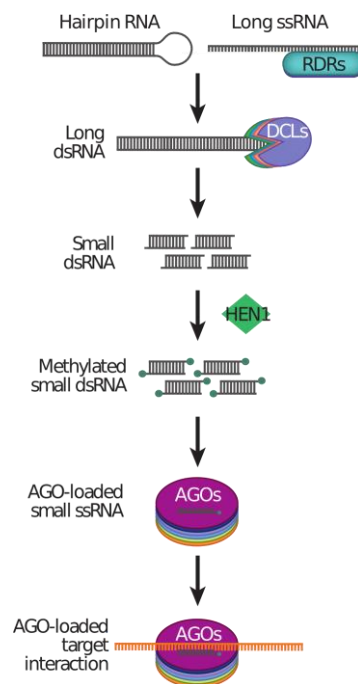
intermediarios más cortos a los que llamaron ARNs pequeños de interferencia (siARN por sus siglas en inglés), y que eran capaces de unirse a ARNs complementarios llevando a la degradación del transcripto blanco (Hammond et al. 2000; Zamore et al. 2000). El descubrimiento de las moléculas de ARN pequeñas, con capacidad de regular genes e incluso genomas, ha sido uno de los avances más significativos en biología en las últimas décadas. Estas moléculas son ARN no codificantes de aproximadamente 20-30 nucleótidos de longitud que actúan como guías para la regulación de múltiples procesos biológicos. Los efectos de estos ARNs pequeños sobre el control de la expresión de genes es generalmente inhibitorio y el mecanismo regulatorio correspondiente se incluye dentro de los mecanismos de silenciamiento génico.

### 1.1.2. Silenciamiento génico post-transcripcional por ARNs pequeños en plantas

Hasta el momento se conocen siete clases de ARNs pequeños reguladores de la expresión génica en plantas, agrupados en dos subclases según el nivel al cual actúan: los de regulación transcripcional y los de regulación post-transcripcional (Bologna & Voinnet, 2014a). Los ARNs pequeños de regulación post-transcripcional se dividen en cuatro grupos según aspectos de sus orígenes, estructuras, proteínas efectoras asociadas y funciones biológicas: (i) los micro-ARNs (miARNs), (ii) los ARNs pequeños de interferencia derivados de secuencias repetidas invertidas (siARNs) (iii) los ARNs pequeños de interferencia derivados de transcritos antisentido naturales (nat-siARNs), y (iv) los ARNs pequeños de interferencia de acción en trans (tasiARNs). En todos los casos, la identidad de los genes que van a ser silenciados está especificada por la secuencia del ARN pequeño, que es capaz de reconocer a su blanco por complementariedad de bases. Los mecanismos de silenciamiento génico mediado por ARNs pequeños comparten cuatro pasos conservados y necesarios (**Error! No se encuentra el origen de la referencia.**): (i) el procesamiento de un ARNdH largo para liberar un ARNdH pequeño, de un tamaño de 18-25 pares de bases (pb) de longitud, (ii) la metilación de ambos extremos 3' del ARNdH pequeño, (iii) la selección de una de las hebras del ARNdH pequeño y su incorporación al complejo efector que contiene una proteína de la familia ARGONAUTA (AGO) y, por último, (iv) el reconocimiento del ARN blanco por complementariedad de bases y su silenciamiento (Chapman & Carrington 2007).

El origen de la molécula de ARNdH larga que va a ser procesada es diverso. La misma puede derivar de la replicación de un virus, de la transcripción de secuencias invertidas, de estructuras de tallo y bucle intramoleculares de un transcripto o de la síntesis de un ARNdH por acción de polimerasas de ARN dependientes de ARN (RDR, del inglés *RNA-Dependent RNA polymerase*).

El procesamiento del ARNdH involucra la participación de ribonucleasas clase III denominadas Dicer (Bologna & Voinnet 2014a). Las proteínas Dicer se encuentran altamente conservadas en las células eucariotas y están presentes tanto en organismos unicelulares como en plantas superiores y vertebrados. En plantas las enzimas homólogas a Dicer se denominan DCL (del inglés *Dicer-Like*). Luego de que DCL produce la liberación del dúplex miARN/miARN\* o siARN/siARN\* esta molécula es protegida de la degradación por acción de la enzima ARN metil transferasa HUA ENHANCER 1 (HEN1), que introduce un grupo metilo en el grupo 2'OH de la ribosa en cada extremo 3' del dúplex (Boutet et al. 2003). A continuación, esta molécula es transportada al citoplasma por el transportador HASTY (HST) (Park et al., 2005). Una vez en el citoplasma, una hebra del dúplex se asocia con AGO y sirve como guía para el complejo RISC (por sus siglas en inglés *RNA induced silencing complex*) (Llave et al. 2002). Este complejo, lleva a cabo la búsqueda de ácidos nucleicos complementarios al ARN pequeño simple hebra para ejecutar su función inhibitoria.



**Figura 1.1 Silenciamiento génico mediado por ARNs pequeños en plantas.**

El silenciamiento génico mediado por ARNs pequeños consta de cuatro pasos básicos: procesamiento de un ARNdH largo para liberar el ARNdH pequeño, metilación de los extremos 3' de los ARNdH pequeños, selección de una de las hebras e incorporación en el complejo con AGO, reconocimiento del ARN blanco por complementariedad de bases y silenciamiento post-transcripcional (Bologna & Voinnet, 2014a).

La función inhibitoria post-transcripcional puede darse por dos mecanismos: (i) corte endonucleolítico sobre ARN mensajeros blancos, o (ii) represión de la traducción de los mismos (silenciamiento post-transcripcional) (Bologna & Voinnet, 2014a). La función efectora que se lleve a cabo, va estar definida por la identidad de la proteína AGO que forme parte del complejo RISC. En plantas, existe redundancia en los componentes de la maquinaria de silenciamiento por ARNs pequeños. La planta modelo *Arabidopsis thaliana* cuenta con diez proteínas AGO y seis polimerasas dependientes de ARN (RDRs). Todas ellas están altamente conservadas y colaboran en varias etapas dentro de las diferentes rutas de silenciamiento. El modo en que cada clase de ARN pequeño es dirigido hacia una única vía es, hasta el momento, poco conocido.

## 1.2. Biogénesis de ARNs pequeños en plantas

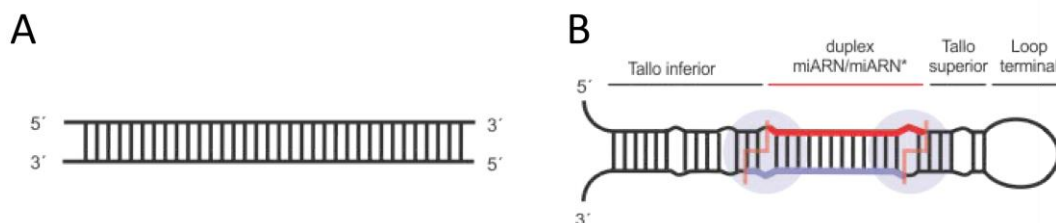
Las diferencias fundamentales en la biogénesis de los distintos tipos de ARNs pequeños se encuentran en el tipo de precursor que les da origen y las proteínas involucradas en la vía de procesamiento. En *A. thaliana* existen cuatro proteínas homólogas a Dicer denominadas DCL1 a 4. Cada una de estas enzimas participa en la biogénesis de distintos ARNs pequeños.

### 1.2.1. siARNs, tasiARNs y nat-siARNs

Los siARNs representan a los inductores del silenciamiento por ARNi canónicos. Sus precursores son moléculas de ARNdH largas, con complementariedad de bases perfecta y de estructura lineal (Figura 1.2, A). Dichos precursores, pueden ser introducidos directamente desde el medio ambiente o pueden ser producidos endógenamente por las células. Los siARNs fueron inicialmente identificados durante el silenciamiento génico inducido por transposones y virus en plantas (Mello & Conte 2004). Una vez que estas moléculas de ARNdH son detectadas en el



citoplasma de la célula, se induce la respuesta de defensa. La respuesta consiste en el procesamiento solapado por DCL2 o DCL4, generando siARNs de 21-22 nt, y DCL3, generando siARNs de 24 nt, que serán los efectores del silenciamiento post-transcripcional (Figura 1.3, A) (Bologna & Voinnet, 2014a; Xie & Qi, 2008). DCL1 estimula la acumulación de siARNs, aparentemente por mediar la separación del tallo doble hebra perfecto de la región simple hebra de los transcritos de repeticiones invertidas. La pérdida de función de la proteína accesoria de DCL4, DRB4, provoca un fuerte aumento en la producción de siARNs de 24 nt dependientes de DCL3. Esta vía no muestra dependencia con las ARN polimerasas dependientes de ARN (RDRs), consistente con la naturaleza de las regiones repetidas invertidas que dan origen a los precursores (Bologna & Voinnet 2014a).



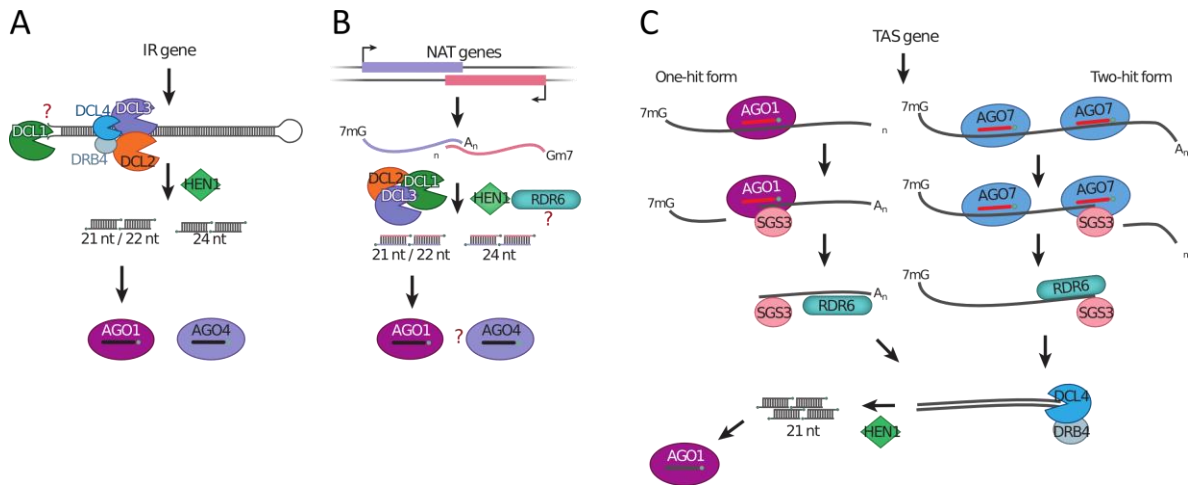
**Figura 1.2 Estructuras de precursores de ARNs pequeños en plantas.**

A. Estructura básica de precursores de ARNs pequeños de interferencia (siARNs, tasiARNs, nat-siARNs). B. Estructura básica de precursores de miARNs.

El grupo de los siARN naturales antisentido (nat-siARN) se origina a partir de ARNdh formados por la transcripción bidireccional convergente de dos genes parcialmente solapados (Borsani et al. 2005; Katiyar-Agarwal et al. 2006). En los casos reportados hasta el momento siempre uno de los genes solapados presenta una expresión constitutiva. La inducción de la expresión del gen complementario activa la formación del ARNdh, el cual es procesado por DCL1 o DCL2 liberando nat-siARN primarios (Figura 1.3, B). Estos siARN primarios son capaces de inducir el corte sobre el ARNm expresado constitutivamente. Existen evidencias de que a partir de este ARNm una RDR podría sintetizar una hebra complementaria. Esto generaría una molécula de ARNdh que podría ser procesada por DCL1 o DCL3 para producir una sucesión de siARNs denominados nat-siARN secundarios (Borsani et al. 2005; Katiyar-Agarwal et al. 2006). Esta última parte de la vía es aún poco conocida y requiere mayor investigación (Bologna & Voinnet 2014a).

Los siARN de acción en trans (tasiARNs) se descubrieron a partir de estudios funcionales en plantas. La biogénesis de estas moléculas de ARN pequeñas es especialmente compleja (Figura 1.3, C). Los genes que dan lugar a los tasiARNs son transcritos por la ARN polimerasa II, a partir de genes TAS, dando lugar a un ARN no codificante. Hasta el momento se conocen dos vías de biogénesis de tasiARNs a partir de este transcrito. Una de las vías requiere el reconocimiento del transcrito por un miARN en complejo con AGO1 (Figura 1.3, C, izquierda), mientras que en la otra el transcrito es reconocido en dos sitios distintos por miARNs en complejo con AGO7 (Figura 1.3, C, derecha). En ambos casos la proteína AGO realiza un corte de manera tal que el miARN actúa como cebador para la transcripción de ARNdh por parte de una RDR específica. Finalmente el ARNdh generado es procesado por DCL4, que realiza cortes sucesivos secuenciales, liberando tasiARNs de 21 nts cuya secuencia está especificada por la fase generada a partir del corte del miARN inicial. Algunos de los ARNs pequeños formados son

incorporados al complejo RISC y guían la degradación de ARNm complementarios (Allen & Howell 2010; Peragine et al. 2004; Montgomery et al. 2008; Cuperus et al. 2010).



**Figura 1.3 Vías de biogénesis de ARNs pequeños de interferencia en *Arabidopsis thaliana*.**

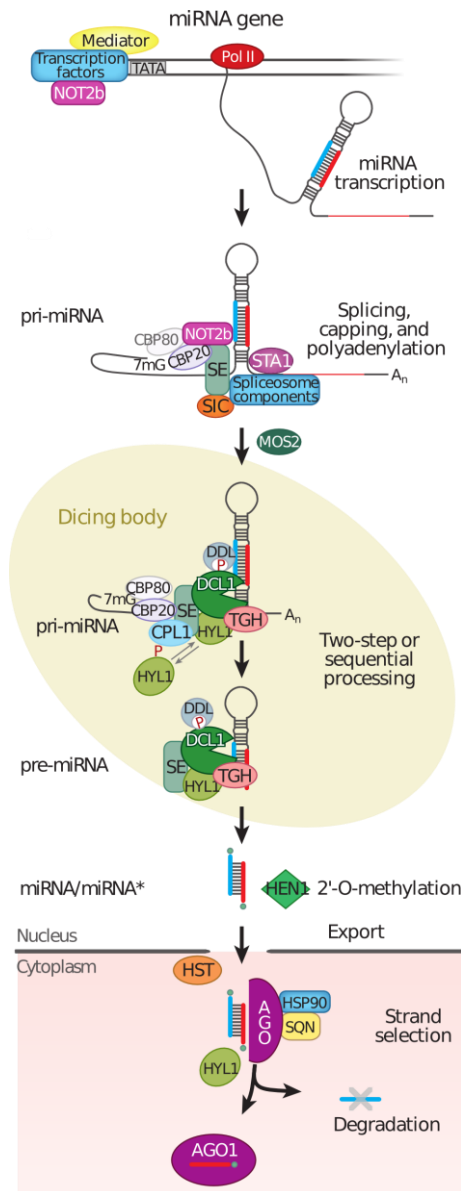
A. Vía de siARNs derivados de secuencias invertidas repetitivas. B. Vía de siARNs derivados de transcritos antisentido naturales. C. Vía de siARNs de acción en trans (Bologna & Voinnet, 2014a).

### 1.2.2. micro-ARNs

Los miARNs se procesan a partir de precursores primarios más largos (pri-miARNs) que son transcritos en el núcleo por la ARN polimerasa II. A diferencia de los precursores de siARNs, los pri-miARN forman estructuras en forma de hebilla, con un tallo doble hebra imperfecto y un bucle terminal que posteriormente debe ser procesado para obtener la molécula efectora de 21 nt de longitud (Figura 1.2, B). Durante este último tiempo, se han identificado miles de miARNs en diversos animales y plantas, al principio mediante clonado de sus secuencias y, posteriormente, a través de análisis computacionales de genomas completos. Al presente hay anotados 28.645 miARNs (miRBase, septiembre de 2015). De los miles de miARNs identificados, sólo se conocen las funciones biológicas de unos pocos. En el caso de los miARN de plantas se ha observado que, generalmente, presentan un alto grado de complementariedad con el ARNm blanco y que su principal función es inducir la digestión del mismo (Llave et al. 2002). Sin embargo, en algunos casos se ha reportado que estas moléculas cumplen la función de inhibir la síntesis de proteínas (Aukerman & Sakai 2003; Chen 2004). Una característica particular de los blancos de miARNs de plantas conocidos hasta el momento es que la mayoría son factores de transcripción, y que están principalmente involucrados en la regulación del desarrollo o de la diferenciación celular. En contraste, los ARNms blanco de los miARNs de animales están relacionados, generalmente, con funciones mucho más diversas. Otra característica destacable de estas moléculas es que, tanto en animales como en plantas, una sola especie de miARN puede unirse a diferentes ARNms blanco y, además, muchos miARNs distintos pueden cooperar en el control de un mismo ARNm. Esta posibilidad genera una red compleja de regulación que asegura el funcionamiento correcto de los distintos procesos biológicos (Kim 2005).

El genoma de *Arabidopsis* contiene cientos de regiones que codifican para miARNs denominadas genes MIR. Al igual que los genes codificantes para proteínas, los genes MIR se transcriben por la ARN polimerasa II y los transcritos primarios sufren el agregado de la caperuza en el extremo 5' y la poliadenilación en el extremo 3' (Xie et al. 2005). También son

blanco de *splicing* alternativo, particularmente en la región 3' adyacente a la estructura de hebilla (Bologna & Voinnet, 2014a).



**Figura 1.4 Vía de biogénesis de miARNs en *Arabidopsis thaliana*.**

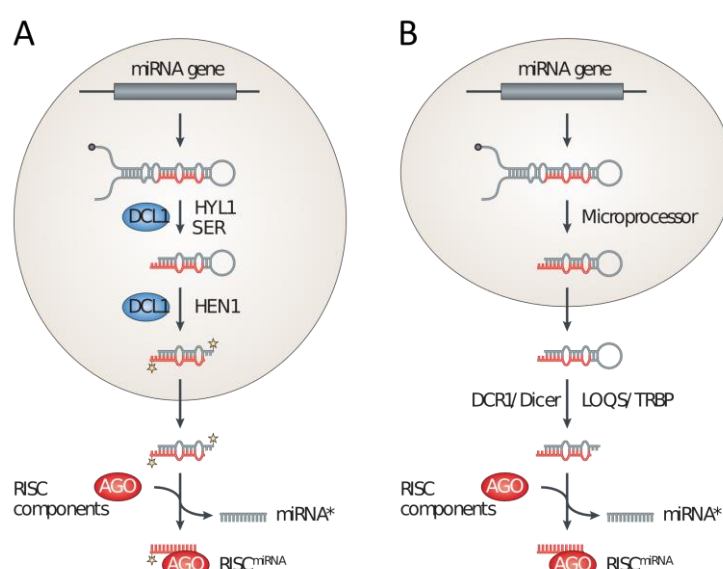
Dentro del núcleo se transcribe el pri-miARN, se produce la maduración del transcrito, se dirige al cuerpo D, se procesa liberando el dúplex miARN/miARN\*, se metila y se exporta al citoplasma. En el citoplasma se selecciona una de las hebras para formar el complejo con AGO y llevar a cabo la función reguladora (Bologna & Voinnet, 2014a).

Una vez ocurrido el *splicing*, el transcrito es direccionado a los sitios específicos dentro del núcleo donde se concentra el procesamiento de pri-miARNs, los denominados cuerpos D (del inglés *Dicing-Bodies*) (Bologna & Voinnet, 2014a) (Figura 1.4). Ya en los cuerpos D, el precursor del miARN sufre al menos dos cortes por parte de DCL1. El procesamiento ocurre íntegramente dentro del núcleo (Park et al. 2002). El primer corte que realiza DCL1 sobre el pri-miARN genera un precursor plegado más corto, denominado pre-miARN, con dos nucleótidos protuberantes en el extremo 3'. A continuación, el pre-miARN es cortado nuevamente por la misma enzima para liberar el dúplex miARN/miARN\* (Kurihara & Watanabe 2004). DCL1 actúa

en conjunto con las proteínas accesorias HYPONASTIC LEAVES 1 (HYL1), SERRATE (SE) y componentes del complejo de unión a la caperuza (CBC) (Gregory et al., 2008; Laubinger et al., 2008; Lobbes et al., 2006; Yang et al., 2006). La fosfatasa CPL1, que co-localiza con las proteínas DCL1 y HYL1, participa en esta vía de una manera indirecta activando a HYL1 por desfosforilación de residuos específicos (Manavella et al. 2012). Por otra parte, mutaciones en DAWDLE (DW) (Yu et al. 2008), TOUGH (TGH) (Ren et al. 2012), SICKLE (SIC) (Zhan et al. 2012) y MOS2 (Wu et al. 2013) también afectan la biogénesis de miARNs.

Uno de los puntos a comparar entre la biogénesis de los ARN pequeños del apartado anterior y la de los miARNs es la forma en que se define la secuencia del ARN pequeño maduro. Para generar siARNs, tasiARNs y nat-siARNs una enzima DCL digiere un precursor completo de ARNdH largo en fragmentos consecutivos de 21 – 24 nt, por lo tanto la secuencia del ARN pequeño sólo depende de la secuencia del precursor y el registro en el que actúe la enzima que lo procesa. Este no es el caso para los miARN, en el cual el fragmento liberado a partir del precursor más largo es específicamente seleccionado por la maquinaria de procesamiento.

### 1.3. Biogénesis de miARNs en animales vs. plantas



**Figura 1.5 Diferencias en la biogénesis de miARN en plantas y animales.**

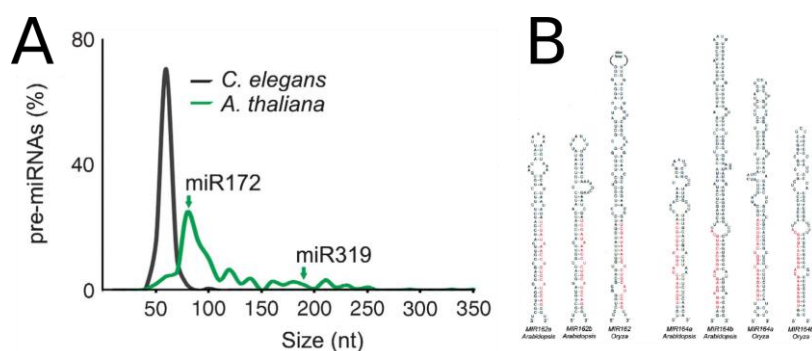
*A. Biogénesis de miARNs en plantas. B. Biogénesis de miARNs en animales (Chapman & Carrington 2007).*

A pesar de tratarse de mecanismos de silenciamiento génico homólogos, las vías de producción de miARNs presentan grandes diferencias entre animales y plantas. La biogénesis de miARNs en animales ha sido extensamente caracterizada. Una de las grandes diferencias con las plantas radica en que en animales el procesamiento de los precursores ocurre en dos pasos separados en tiempo y espacio (Figura 1.5, B), mientras que en plantas ocurre íntegramente en el núcleo (Figura 1.5, A). En animales la primera etapa del procesamiento ocurre en el núcleo. El complejo microprocesador formado por la ARNasa III Drosha y una proteína de unión a ARNdH (DGCR8 en humanos y PASHA en *D. melanogaster*) realiza el primer corte del pri-miARN liberando un precursor más corto, el pre-miARN. En el complejo microprocesador también participan una variedad de cofactores que incluyen la helicasa de ARN p68 (DDX5) y la p72 (DDX17), así como

ribonucleoproteínas heterogéneas nucleares (hnRNPs) (Gregory et al. 2004). Estas proteínas auxiliares, pueden tener funciones en la fidelidad, especificidad y/o en la actividad del corte que realiza Drosha. Drosha realiza el corte de manera concertada con la transcripción y genera un producto típico de enzimas RNasa III, con dos nucleótidos protuberantes en el extremo 3' de la molécula (el mecanismo detallado del reconocimiento del sustrato por parte del complejo Drosha/DGCR8 se detalla en el apartado 1.5.2). Este extremo característico permite que el pre-miARN sea reconocido por *Exportin 5*, que transporta el sustrato hacia el citoplasma (Kim et al. 2009). Una vez en el citoplasma, la RNasa III Dicer corta los primeros 22 nucleótidos del pre-miARN con la participación de las proteínas de unión a ARNdh accesorias TRBP (proteína de unión al ARN TAR) y PACT (proteína activadora de PKR) en mamíferos (Loquacious (LOQS) y R2D2 en *D. melanogaster*) (Chendrimada et al., 2005; Förstemann et al., 2005; Haase et al., 2005; Kim et al., 2009; Kok et al., 2007; Lee et al., 2006; Saito et al., 2005). Posteriormente, las dos hebras del miARN son separadas y una de ellas se asocia con la proteína AGO que forma parte del complejo RISC. Una vez en el complejo de silenciamiento, el miARN sirve como guía hacia sitios complementarios dentro de ARNs blanco para mediar la represión de los mismos.

#### 1.4. Determinantes estructurales en precursores de miARN

Además de las diferencias en los pasos de las vías de biogénesis, otra de las grandes diferencias entre plantas y animales consiste en las estructuras de los precursores de miARNs. Los precursores de animales se caracterizan por presentar una estructura relativamente homogénea, que consta de un tallo imperfecto de una longitud de *ca.* 65 nt, flanqueado por dos regiones desapareadas. Estas dos regiones son: un bucle terminal en un extremo y ARNsh en el otro. De esta manera, los pri-miARNs de animales pueden ser divididos en cinco partes: un bucle terminal, un dúplex miARN/miARN\* (dos vueltas de ARNdh), un tallo inferior de una vuelta de ARNdh y secuencias flanqueantes simple hebra. En animales, la especificidad en la selección de la secuencia del miARN dentro del precursor está dada por el primer corte realizado por el complejo microprocesador. Por otro lado, como se muestra en la Figura 1.6, los precursores de miARNs de plantas presentan tamaños que varían desde los 50 a 900 nt, lo que se traduce en un amplio abanico de estructuras diferentes (Bologna et al. 2009; Chu & Rana 2007).

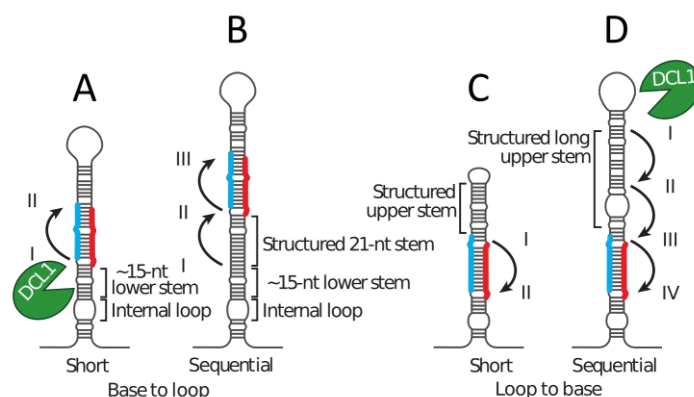


**Figura 1.6 Los precursores de miARN de plantas presentan mayor heterogeneidad que los de animales.**

A. Heterogeneidad de los precursores de miARNs en plantas en términos de longitud respecto a los de animales. B. Heterogeneidad en la estructura de la hebra y la posición del miARN. (Bologna et al. 2009)

A la heterogeneidad en estructura y secuencia se suma que en plantas no todos los precursores se procesan a través del mismo mecanismo. Es decir, al contrario de lo que sucede en animales, los pri-miARNs en plantas pueden sufrir el primer corte desde ambos extremos de la hebra,

tanto desde el bucle como desde la base. Además, hay precursores largos que requieren más de dos eventos de corte sucesivos para que se produzca la liberación del miARN maduro (Figura 1.7). Esto ha permitido agrupar los precursores según su mecanismo de procesamiento. Analizando las secuencias y estructuras de los pri-miARN dentro de cada grupo se han identificado ciertas características comunes que podrían estar indicando el sitio de corte a la maquinaria de procesamiento.



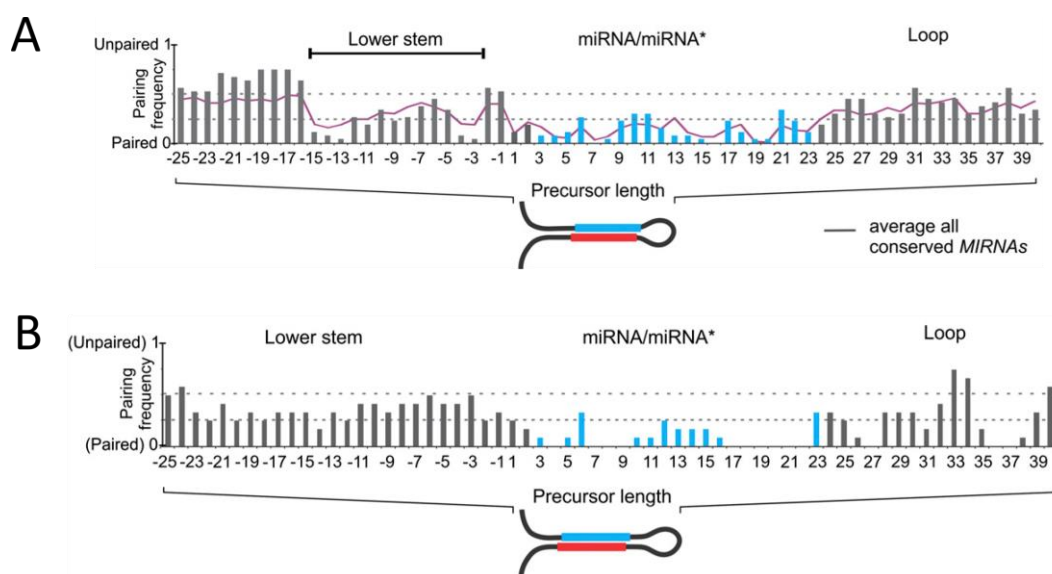
**Figura 1.7 Los precursores en plantas se procesan por cuatro mecanismos diferentes.**

*A. Procesamiento corto desde la base hacia el bucle. B. Procesamiento largo desde la base hacia el bucle. C. Procesamiento corto desde el bucle hacia la base. D. Procesamiento largo desde el bucle hacia la base (Bologna & Voinnet 2014a).*

El primer mecanismo de procesamiento se denomina mecanismo corto desde la base hacia el bucle (Figura 1.7, A). Este mecanismo se da en precursores que contienen una región desapareada, denominada bucle interno, seguida de un tallo inferior de entre 15 y 17 nt de longitud parcialmente desapareados ubicado debajo de la secuencia del miARN/miARN\*. El tallo estructurado de ca. 15 nt ha sido definido como el determinante estructural que permite a la maquinaria de procesamiento localizar el sitio del primer corte (Bologna et al., 2013; Zhu et al., 2013; Mateos et al., 2010; Song et al., 2010; Werner et al., 2010). Este determinante también está presente en precursores más largos que son procesados a través del mecanismo largo desde la base hacia el bucle (Figura 1.7, B). En estos casos el sitio del primer corte es definido por el tallo inferior, mientras que el miARN maduro es liberado luego de tres cortes sucesivos, produciendo como subproducto un miARN intermedio que se acumula en baja cantidad en las células (Bologna et al., 2013; Kurihara & Watanabe, 2004).

Los restantes dos mecanismos se producen en la dirección opuesta. El primero en ser identificado fue el largo desde el bucle hacia la base (Figura 1.7, D). Los precursores que son procesados a través de este mecanismo presentan un tallo largo desde el miARN/miARN\* hacia el bucle terminal, denominado tallo superior, que es esencial para el procesamiento (Addo-Quaye et al., 2009; Bologna et al., 2009). La maquinaria de procesamiento comienza realizando un corte cercano al bucle terminal y luego lleva a cabo una serie de cortes sucesivos separados por 21 nt hasta liberar el miARN/miARN\* maduro. Más recientemente han sido identificados una serie de precursores que son procesados en esta misma dirección pero que sólo requieren dos cortes para que el miARN sea liberado. Este mecanismo se denomina corto desde el bucle hacia la base (Figura 1.7, C). Estos precursores comparten un tallo estructurado de ca. 42 nt de longitud total, incluyendo el dúplex miARN/miARN\*, y un bucle pequeño (Bologna et al., 2013).

En un estudio masivo reciente se confirmaron los mecanismos de procesamiento ya conocidos, se definieron los mecanismos para precursores nuevos y se describió el mecanismo corto desde el bucle hacia la base (Bologna et al. 2013). La metodología empleada para identificar los determinantes estructurales, que definen el mecanismo a través del cual es procesado un precursor, se basa en estructuras calculadas *in silico* empleando el servidor Mfold (Zuker 2003). En el trabajo realizado por Palatnik y colaboradores (Bologna et al. 2013) se agrupan los precursores procesados a través de un mismo mecanismo y se busca alinear sus estructuras secundarias calculadas. Comparando el grado de apareamiento de las distintas regiones los autores definen el tallo inferior de 15 pb necesario para los procesamientos desde la base hacia el bucle (Figura 1.8, A). En el caso de los procesados desde el bucle a la base observan un grado de apareamiento conservado en el tallo superior pero no identifican una longitud tan definida (Figura 1.8, B).



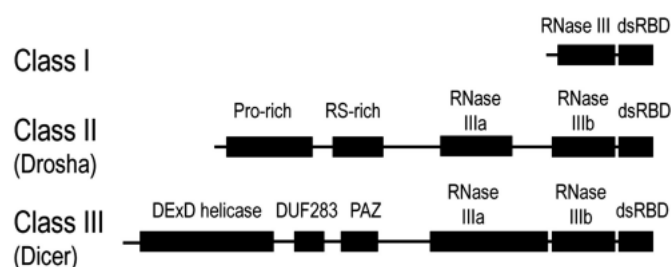
**Figura 1.8** Determinantes estructurales que definen el mecanismo de procesamiento de los pri-miARN.

A. Frecuencia de apareamiento promedio para precursores procesados desde la base hacia el bucle, se destaca la región del tallo inferior de 15 pb (lower stem) y el bucle mayormente desapareado (loop). B. Frecuencia de apareamiento promedio para precursores procesados desde el bucle hacia la base, se destaca que la región mayormente apareada es la del tallo superior comprendida entre el miARN/miARN\* y el bucle (loop). En cada posición se asignó un valor de 1 para nucleótidos desapareados y 0 para nucleótidos apareados (Bologna et al. 2013).

## 1.5. Enzimas ARNasa III, bases estructurales del reconocimiento del sustrato

Como vimos, en plantas, una única enzima (DCL1) realiza todos los cortes necesarios sobre los precursores para liberar el miARN maduro. Esto implica que, de alguna manera, esta enzima está reconociendo el sitio del primer corte específicamente dentro de la gran heterogeneidad estructural de los pri-miARNs. Esta capacidad la diferencia de Dicer de animales, la cual recibe los precursores pre-procesados por Drosha. Por esta razón resulta de interés revisar las características estructurales de las enzimas ARNasa III y la forma en que reconocen el sustrato. Nuestra intención es la de encontrar indicios sobre las particularidades que podrían estar confiriendo a DCL1 la capacidad de reconocer el sitio del primer corte.





**Figura 1.9 Clases de enzimas ARNasa III.**

*Clase I: ARNasas III procariotas, Clase II: ARNasas III tipo Drosha, Clase III: ARNasas III tipo Dicer* (Du et al. 2008).

Las enzimas ARNasa III son específicas de ARNdh. Los productos de la digestión por ARNasa III son fragmentos de ARNdh con un grupo fosfato 5' y un hidroxilo terminal 3' en cada hebra. Además presentan 2 nt protuberantes simple hebra en el extremo 3'. Las distintas enzimas ARNasa III presentan dominios adicionales que funcionan concertadamente con los dominios ARNasa III. Estos dominios adicionales contribuyen a la unión del sustrato, amplían el rango de los sustratos posibles, participan en la selección del sitio de corte, confieren elementos regulatorios y posibilitan interacciones proteína–proteína (Nicholson 2014). Según el arreglo de dominios, este tipo de enzimas puede dividirse en tres clases (Figura 1.9) (MacRae & Doudna 2007). Las de clase I se encuentran en bacterias, bacteriófagos y hongos, y contienen un solo dominio tipo ARNasa III unido a un dominio de unión a ARNdh (dsRBD). Las representantes de las clases II y III son Drosha y Dicer, respectivamente. Las de clase III son las más complejas y poseen en general un dominio helicasa (DExD/H-box), un pequeño dominio de interacción proteína–proteína (DUF283), un dominio PAZ (Piwi Argonaut Zwill), dos dominios ARNasa III en tándem, y al menos un dsRBD.

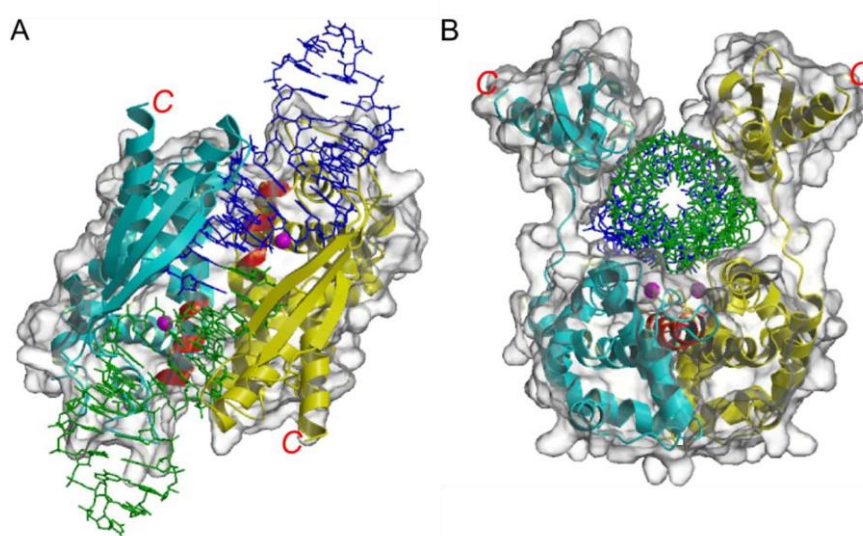
El mecanismo propuesto de acción de estas enzimas consiste en la dimerización inter o intramolecular de los dominios ARNasa III para formar un centro catalítico único (Zhang et al. 2004). En el caso de las enzimas de clase I el dímero se forma entre dos moléculas distintas, mientras que para las de clase II y III se forma un pseudo-dímero intramolecular. En este centro catalítico único cada uno de los dominios ARNasa corta una de las hebras del ARNdh. La distancia entre los sitios catalíticos de los dos dominios involucrados es la que genera el típico extremo 3' con 2 nt protuberantes.

#### 1.5.1. Clase I: ARNasa III

Las ARNasas III bacterianas consisten típicamente de un dominio ARNasa III, un conector flexible y un dominio de unión a ARNdh C-terminal. La enzima más estudiada de este grupo es la ARNasa III de *Escherichia coli*. Esta enzima juega un rol importante en el procesamiento de precursores de ARNs ribosomales y regula la traducción actuando sobre ARNs mensajeros blanco. Los sustratos típicos de esta enzima son ARNs propios de la célula o virales que poseen dos segmentos complementarios capaces de hibridar entre sí y formar regiones de ARNdh. La combinación de estudios bioquímicos y estructurales sobre distintas enzimas de esta clase pertenecientes a diferentes especies ha permitido definir una serie de características de funcionamiento comunes. En este sentido una de las estructuras resueltas más interesantes fue determinada por Ji y colaboradores para la ARNasa III de *Aquifex aeolicus* (Aa-ARNasa III) (Gan et al. 2006). Empleando cristalografía de difracción de rayos X, lograron determinar la estructura de la enzima en presencia del producto de la reacción a una resolución de 2,05 Å. La



estructura confirma la formación de un dímero intermolecular sobre el sustrato (Figura 1.10) que ya había sido sugerida por evidencias bioquímicas anteriores. A partir de esta estructura los autores determinaron una serie de características que afectan la unión con el ARNdh. Comparando la estructura en presencia de ARNdh con otras estructuras determinadas en ausencia del mismo observan que el dominio dsRBD adquiere una posición definida por la unión al ARN, mientras que en las formas libres de las distintas enzimas los dominios dsRBDs no presentan una posición definida respecto del dominio ARNasa III. Esto se debe a la presencia de un conector de longitud variable entre los dominios. La flexibilidad de este conector es esencial para la actividad catalítica de la enzima, como fue demostrado para la enzima ARNasa III de *E. coli* en un trabajo anterior (Inada & Nakamura 1995). Dos tercios de las interacciones de la proteína con el sustrato ocurren a través del dsRBD.



**Figura 1.10 Estructura cristalográfica de ARNasa III de *Aquifex aeolicus* en complejo con el ARNdh producto**

*En celeste y amarillo se muestran los dos monómeros de la enzima. En verde y azul el ARNdh. A. Los dominios proteicos en el primer plano son los dsRBDs. B. Los dominios en la parte superior son los dsRBDs y en la parte inferior son los ARNasa III. (Gan et al. 2006)*

*In vitro*, la ARNasa III bacteriana es capaz de procesar cualquier ARNdh perfecto, sin ninguna preferencia de secuencia, dando como resultado fragmentos de entre 11 y 15 nucleótidos (Nicholson 2014). Sin embargo, los ARNs blanco de las ARNasa III tanto de bacterias como de bacteriófagos son escindidos en posiciones específicas. A pesar de que los sitios de reconocimiento se han estudiado durante décadas, aún no ha sido posible identificar una secuencia consenso clara. Se han logrado identificar dos regiones importantes sobre el sustrato para su correcto procesamiento. Estos sitios se han denominado caja distal y proximal respecto del sitio de corte. La caja distal consiste en cuatro pares de bases ubicados a aproximadamente once nucleótidos de distancia del sitio de corte (Pertzev & Nicholson 2006). Esta distancia sirve para definir el sustrato de menor longitud necesario para que se produzca una reacción eficiente. Este determinante no presenta una secuencia conservada ni entre los sustratos de una misma ARNasa ni entre diferentes especies (Nicholson 2014). La caja proximal consiste en dos pares de bases adyacentes al sitio de corte y su secuencia controla la reactividad del sustrato y define la selección del sitio blanco. Esta región es contactada por cadenas laterales de

la hélice  $\alpha$  del extremo N-terminal del dsRBD de la enzima (Gan et al. 2008; Gan et al. 2006). Una de las interacciones de mayor interés entre el dsRBD y el ARNdh involucra un puente de hidrógeno entre la cadena lateral de una glutamina altamente conservada en los dsRBDs en general (157Q en Aa-ARNasa III) y el O 2 del anillo pirimidina de un uracilo en un par AU de la caja proximal (Gan et al. 2008; Gan et al. 2006; Lamontagne & Elela 2004). La cadena lateral de este residuo también interacciona con un grupo 2'OH de un residuo de ribosa adyacente. Estas interacciones son de importancia funcional, ya que una mutación de esta glutamina por una alanina reduce la actividad catalítica de la ARNasa III debido a la disminución de la afinidad de la enzima por el sustrato (Shi et al. 2011). La importancia de la interacción entre el dsRBD y la caja proximal se ve reflejada también en la dependencia de la reacción con la secuencia del ARN en esta posición. Mientras un par AU o UA en esta posición confieren reactividades equivalentes, el reemplazo por un par GC o CG bloquean la unión de la enzima, evitando el procesamiento (Pertzev & Nicholson 2006; Shi et al. 2011).

En relación a la problemática de las enzimas ARNasa III en general, los aportes de los estudios de las enzimas bacterianas han sido de gran importancia. El modelo del dímero intermolecular para la conformación de un sitio catalítico único ha sido la base para las hipótesis sobre la formación del sitio activo en enzimas ARNasa III más complejas que presentan ambos dominios ARNasa III sobre la misma molécula, formando un dímero intramolecular. Por otro lado, al demostrarse la dependencia de la eficiencia de la enzima con la unión al sustrato a través del dsRBD, sugiere que estos dominios podrían jugar roles similares en las enzimas eucariotas. Incluso el reconocimiento de secuencia de la caja proximal ocurre a través de este dominio.

#### 1.5.2. Clase II: Drosha

Como vimos, Drosha es la ARNasa III a cargo de realizar el primer paso de procesamiento de precursores de miARNs en animales. Esta enzima por sí sola no es capaz de procesar los precursores. Para ejercer su función necesita la asistencia de la proteína de unión a ARN DGCR8 (Figura 1.11, A) (Han et al. 2004; Han et al. 2006).

Hasta el momento se dispone de relativamente poca información estructural sobre Drosha completa. Se sabe que esta enzima posee dos dominios ARNasa III en tándem (Figura 1.9). A través de experimentos bioquímicos se ha demostrado que estos dominios funcionan en conjunto como un pseudo-dímero intramolecular formando un centro catalítico único (Han et al. 2006; Han et al. 2004), de manera similar a lo propuesto para el dímero formado por la ARNasa III de *E. coli*.

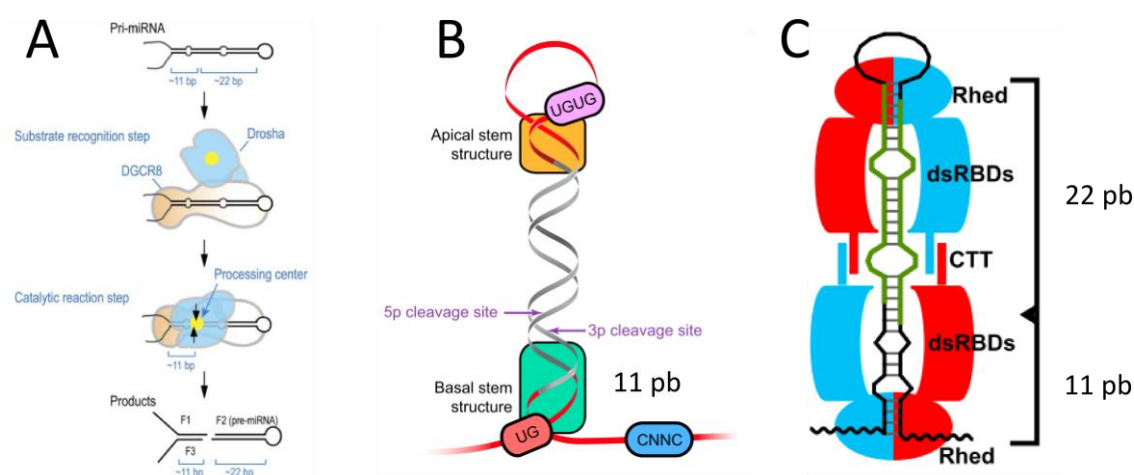
Como las ARNasas III en general, Drosha posee un dsRBD en el extremo C-terminal. Una mutante que presenta una delección en este dominio pierde la función catalítica *in vitro* (Han et al. 2004). Contrariamente a lo esperado, este dominio no presenta afinidad por el ARNdh (Wostenberg et al. 2010). El dominio tampoco es necesario para la interacción de Drosha con DGCR8 (Han et al. 2004) ni posee señal de localización nuclear (Doyle et al. 2013). Por lo tanto la función de este dominio aún resta ser elucidada.

El reconocimiento del sustrato de forma conjunta entre Drosha y DGCR8 resulta en el corte de la doble hebra a una distancia de 11 pb desde la base de la estructura de hebilla de los pri-miARNs (Han et al. 2006). Este corte libera un precursor más corto denominado pre-miARN que presenta un extremo 3' protuberante. Este pre-miARN es el sustrato de Dicer en el citoplasma.

La forma en que los sustratos son reconocidos por el microprocesador para que Drosha realice el corte en el sitio correcto continúa en discusión. En principio los precursores de miARN en

animales presentan una estructura homogénea de hebilla con una región apareada de 11 pb desde el sitio de corte hacia la base del tallo. Esta región doble hebra se encuentra flanqueada por al menos nueve nucleótidos desapareados (Auyeung et al. 2013). Por otro lado, los mismos autores han identificado pequeños determinantes de secuencia que estimularían el reconocimiento por el microprocesador, aumentando la eficiencia en el procesamiento de los precursores (Figura 1.11, B).

DGCR8 posee una señal de localización nuclear, un dominio Rhed (dominio hemo de unión a ARN), dos dsRBDs y un pequeño dominio denominado cola C-terminal (CTT). El último modelo para el reconocimiento de los precursores por DGCR8 es el propuesto por Guo y colaboradores, en el cual cuatro moléculas de DGCR8 se ubican sobre el precursor (Figura 1.11, C) (Quick-Cleveland et al. 2014). Las cuatro unidades forman dos dímeros y reconocen la estructura de hebilla del ARN desde los dos extremos simple hebra. Cada dímero se forma por la interacción entre los dominios Rhed. Al mismo tiempo estos dominios son los responsables del reconocimiento de las regiones simple hebra en la base y en el bucle de la hebilla, cooperando con los dos dsRBDs que reconocen la región doble hebra del precursor. Ambos dímeros interaccionan entre sí a través de la región CTT. Los autores de este modelo proponen que de esta manera DGCR8 realiza un control del tamaño total del precursor, habilitándolo para el procesamiento por Drosha (Quick-Cleveland et al. 2014). Este control del tamaño es posible para precursores de animales dado su tamaño homogéneo (Han et al. 2006; Auyeung et al. 2013).



**Figura 1.11 Reconocimiento del sustrato por parte de Drosha/DGCR8.**

A. Mecanismo propuesto por Kim y colaboradores (Han et al. 2006), DGCR8 es necesaria para la actividad de Drosha. B. Determinantes estructurales y de secuencia definidos por Bartel y colaboradores (Auyeung et al. 2013), el determinante estructural más conservado es la región estructurada en la base del tallo (11 pb). C. Modelo de unión de los precursores por parte de DGCR8 tipo “pinza molecular” (Quick-Cleveland et al. 2014), cuatro subunidades de DGCR8 interactúan con el precursor, un dímero reconoce la transición ARNsh – ARNdh en la base y el otro dímero reconoce la transición en la región del bucle.

### 1.5.3. Clase III: Dicer

La clase III de las ARNasa III agrupa a las enzimas Dicer. Las Dicer son endonucleasas multidominio de gran tamaño que juegan una función central en las vías de miARNs y de ARNi. Estas enzimas se encuentran presentes en la mayoría de los eucariotas, con un número de copias y estructura de dominios que varía entre los distintos organismos. Existen cuatro enzimas

de tipo Dicer en *Arabidopsis thaliana*, dos en *Drosophila melanogaster* y una en mamíferos, *Caenorhabditis elegans*, *Schizosaccharomyces pombe* y en otras especies de levaduras (Park et al. 2005).

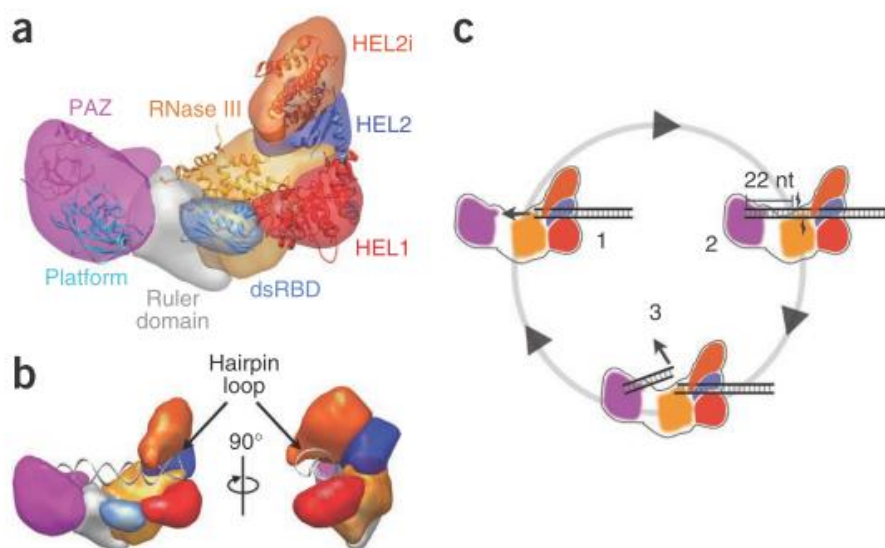
A lo largo de los últimos diez años se han determinado las estructuras de diferentes regiones de enzimas Dicer de varios organismos. La única estructura de alta resolución de una enzima Dicer completa es la del parásito *G. intestinalis* (Macrae et al. 2006) en ausencia del sustrato. Esta enzima es particular entre las Dicer ya que sólo consiste en un dominio PAZ, un pequeño dominio *ruler* y dos ARNasa III. Recientemente se ha determinado la estructura de baja resolución de la enzima Dicer de humanos completa, empleando microscopía electrónica criogénica (Lau et al. 2012). A través de esta técnica se obtiene un volumen dentro del cual se modelan luego las regiones de estructura conocida determinadas a través de otras técnicas. En el caso del modelo para Dicer de humanos de la Figura 1.12 A, se emplearon las estructuras cristalográficas de Dicer de *G. intestinalis* para los dominios PAZ, *ruler* y ARNasa III, el dsRBD corresponde a la estructura cristalográfica del dominio de Dicer de ratón (Du et al. 2008), mientras que los dominios helicasa corresponden a la proteína RIG-I de humanos (un receptor de ARN viral del sistema de respuesta inmune innata) (Kowalinski et al. 2011).

Estudios sobre las enzimas Dicer en animales han revelado diferentes modos de unión al sustrato a través de distintos dominios. Estos modos dan origen a la precisión y especificidad en la biogénesis de ARNs pequeños a partir de los distintos sustratos. Uno de estos modos es a través del anclaje del extremo 3' protuberante de un ARNdH al dominio PAZ. Este dominio está conectado al dominio catalítico de tal forma que el sitio activo se localiza a una distancia fija, actuando como una regla molecular, determinando la longitud del producto de la reacción de corte (Macrae et al. 2006; Lau et al. 2012). La estructura cristalográfica de este dominio junto con una región adyacente denominada plataforma fue resuelta recientemente en complejo con un segmento de ARNdH con extremo 3' protuberante (Tian et al. 2014).

Otro punto de interacción consiste en la unión a regiones de ARNdH a través del dsRBD C-terminal. Esta interacción se da con la misma afinidad por ARNdH perfecto como imperfecto (Wostenberg et al. 2012) y es de gran importancia para la eficiencia de la reacción *in vitro* (Zhang et al. 2004; Ma et al. 2012). Otra función que se ha adjudicado recientemente al dsRBD es la de actuar como señal de localización nuclear. A pesar de que Dicer de animales lleva a cabo su función en el citoplasma de la célula, se ha demostrado que es capaz de entrar y salir del núcleo de forma esporádica. En el mismo trabajo se demostró que la NLS en el dsRBD se encuentra, de alguna manera, oculta por el resto de la enzima y por la interacción con el ARN (Doyle et al. 2013).

El tercer modo de unión requiere el reconocimiento de bucles simple hebra por parte del dominio ATPasa/Helicasa. Por esta cualidad se ha atribuido a este dominio ser el sensor capaz de discriminar pre-miARNs de ARNs doble hebra largos (Tsutsumi et al. 2011). La homogeneidad en el tamaño de los pre-miARN de animales posibilitaría este mecanismo. La actividad hidrolítica de nucleótidos trifosfato de este dominio no es necesaria para la biogénesis de miARNs, lo cual concuerda con la necesidad de realizar un único corte para la liberación del miARN maduro en estos organismos (Figura 1.12, b) (Lau et al. 2012; Taylor et al. 2013). Sin embargo, sí es requerida para la biogénesis de ARNs pequeños de interferencia, para lo cual es necesario que la enzima realice cortes secuenciales sobre un mismo ARNdH largo. De esta forma es posible que la función del dominio helicasa de Dicer sea tipo traslocasa ATP-dependiente, proveyendo la energía necesaria para que se produzcan los múltiples cortes a lo largo de los

sustratos de ARNdh precursores de siARNs (Figura 1.12, c) (Cenik et al. 2011; Welker et al. 2011).



**Figura 1.12** Arquitectura y mecanismo de Dicer humana.

A. mapa de densidad electrónica obtenido por microscopía electrónica criogénica, segmentado y presentando los dominios de estructura conocida ensamblados en su interior. B. modelo de unión a pre-miARN entre el dominio PAZ al extremo 3' y Helicasa al bucle terminal. C. modelo de actividad Dicer para cortes sucesivos sobre un precursor ARNdh largo de siARN. (Lau et al. 2012).

#### 1.5.4. Enzimas tipo Dicer en plantas

A diferencia de lo que sucede en animales, las plantas disponen de al menos cuatro copias de enzimas DCL, involucradas en la biogénesis de distintos ARNs pequeños regulatorios (descrito en la sección 1.2). Al momento, la teoría acerca de la diversificación de estas enzimas en plantas indica que puede haberse originado por la necesidad de estos organismos en adquirir resistencia ante el ataque por virus, bacterias y hongos (Liu et al. 2009).



**Figura 1.13** Arquitectura de dominios de enzimas tipo Dicer en *Arabidopsis thaliana*.

NLS: señal de localización nuclear, DEAD/Helicase: dominio helicasa, Duf283: dominio de función desconocida 283, PAZ: Piwi Argonaut Zwill, RNaseIII: ARNasa III, DRB: dominio de unión a ARNdh (Vazquez 2006).

El arreglo de dominios es el básico de las ARNasa III de clase III (Figura 1.13). Por tratarse de proteínas de gran tamaño, la información estructural disponible sobre estas enzimas es limitada, al igual que para las otras enzimas Dicer. Las funciones de los distintos dominios han sido determinadas combinando evidencias de actividad *in vivo* y ensayos bioquímicos.

Mutaciones en el dominio helicasa, ARNasa III o en los dsRBDs de DCL1 reducen la acumulación de miARN maduros *in vivo* (Kurihara & Watanabe 2004; Willmann et al. 2011). El dominio helicasa es necesario para el correcto procesamiento *in vivo* de al menos algunos pri-miARNs y la actividad de la enzima es dependiente de ATP *in vitro* (Liu et al. 2012; Zhu et al. 2013). El hecho de que son necesarios al menos dos eventos de corte sobre un mismo precursor, el primero para generar el pre-miARN y el segundo para liberar el miARN/miARN\* de 21 nt, posiblemente explica esta dependencia (Bologna & Voinnet 2014a).

El dominio DUF283 presenta una estructura similar a un dsRBD, pero un estudio estructural y bioquímico sobre este dominio de DCL4 muestra un plegamiento no canónico y no se detectó afinidad por el ARNdh. En el mismo trabajo se determinó la capacidad de este dominio de interaccionar con la proteína accesoria de unión a ARN DRB4, mientras que el dominio DUF283 de DCL1 interacciona específicamente con HYL1 (Qin et al. 2010).

Una particularidad que diferencia a DCL1, DCL3 y DCL4 de su contraparte en animales es la presencia de un segundo dsRBD en el extremo C-terminal. A pesar de ser un dominio de unión a ARNdh, el dsRBD C-terminal de DCL1 presenta un plegamiento no canónico y forma complejos heterogéneos con el ARNdh (Burdisso et al. 2012). Una de las funciones atribuidas a este dominio es la de actuar como señal de localización nuclear (Burdisso, datos no publicados). Esta capacidad ha sido descrita para los dominios dsRBD C-terminal de Dicer de humanos, de *C. elegans* y una de las enzimas Dicer de *D. melanogaster* (Dcr-1) (Doyle et al. 2013). Una vez dentro del núcleo estaría involucrado en la localización de DCL1 en los cuerpos D y en interacciones proteína-proteína (Liu et al. 2013).

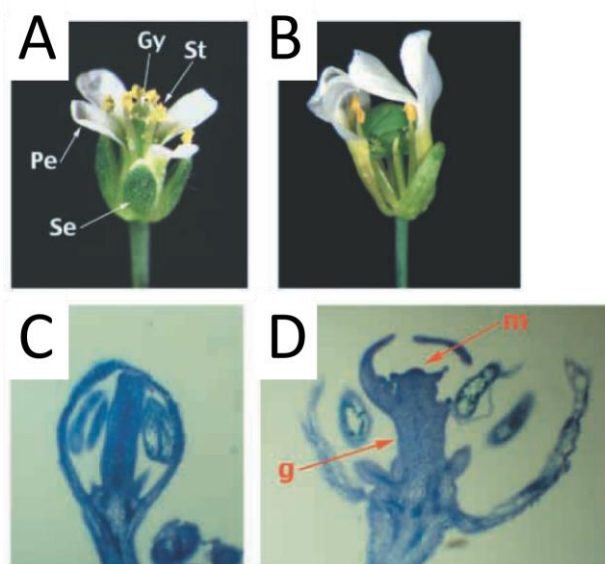
Un estudio filogenético de la similitud entre las secuencias de las cuatro enzimas entre distintas especies de plantas concluye que la diversificación de DCL1 respecto de las otras tres es mayor en comparación con la diversificación de DCL2, 3 y 4 entre sí (Liu et al. 2009). Esto se condice con su función especializada en la vía de biogénesis de los miARN. La pérdida de función de DCL1 resulta letal a nivel de embrión en *A. thaliana*, lo que sugiere que no existe otra DCL que sea capaz de procesar eficientemente a los precursores de miARNs con estructuras de hebilla o de localizarse en el tejido adecuado al momento de cumplir su función esencial (Schauer et al. 2002; Nodine & Bartel 2010).

## 1.6. Relevancia de los dsRBDs en la biogénesis de miARNs

Teniendo en cuenta que las estructuras y funciones de los dominios PAZ, DUF283, ARNasa III han sido estudiadas en profundidad para las enzimas ARNasa III en general y considerando que la diferencia más llamativa en arquitectura de dominios de Dicer entre animales y plantas es la duplicación del dsRBD C-terminal, procederemos a la revisión de evidencia sobre la relevancia de esta región en el procesamiento de miARNs en plantas.

Una característica común entre las proteínas que participan en el procesamiento de los precursores de miARNs es la capacidad de unir ARN. En conjunto, las proteínas que forman el complejo de procesamiento de plantas (DCL1, HYL1 y SERRATE) poseen seis dominios de unión a ARN. Cuatro de ellos pertenecen a la familia de los dsRBDs, que se caracterizan por reconocer específicamente variantes de motivos de ARN doble hebra (Tian et al. 2004). Dos de ellos se encuentran en DCL1 y otros dos son parte de la proteína accesoria HYL1. En este sentido, los dominios de unión a ARNdh cumplirían un rol importante tanto para el reclutamiento del sustrato como para el posicionamiento correcto de los sitios activos.

Los dsRBD son, en principio, capaces de unir cualquier elemento doble hebra en una molécula de ARN. Sin embargo, la presencia de imperfecciones en el ARNdh o la interacción con otras proteínas u otros dominios le permiten tener especificidad sobre las regiones que reconocen (Tian et al. 2004). Esta característica resulta especialmente apropiada para el reconocimiento de precursores de miARN debido a que, como describimos previamente, estos no comparten similitudes en su secuencia y presentan un alto nivel de heterogeneidad. Si bien hasta el momento no se ha establecido una función clara de estos dominios durante la biogénesis de miARNs, se ha demostrado que algunos dominios dsRBD presentan actividad chaperona de ARN, facilitando el apareamiento correcto de segmentos de ARN o promoviendo la formación de estructuras doble hebra en fragmentos que contienen protuberancias (Hitti et al. 1998; Zheng & Bevilacqua 2000). Por esta razón, una función posible para algunos de los dsRBDs involucrados en el procesamiento de miARNs en plantas podría ser la de estabilizar elementos de estructura secundaria para permitir el posicionamiento correcto de los sitios activos de DCL1 en el complejo de procesamiento. Esta interacción daría lugar a la formación de posibles determinantes de la especificidad en la estructura de los precursores que no son obvias en las predicciones de estructura secundaria realizadas por programas bioinformáticos estándar, y permitiría definir sin ambigüedades el registro del procesamiento.



**Figura 1.14 Fenotipo de mutantes *dcl1-9* de *Arabidopsis thaliana*.**

A. Flor de *A. thaliana* silvestre. B. Flor de planta mutante *dcl1-9*. C. Sección transversal de una flor silvestre teñida con azul de toluidina. D. Sección transversal de una flor de planta mutantes *dcl1-9* teñida con azul de toluidina (Jacobsen et al. 1999).

De la sección anterior destacamos que los dsRBDs juegan un rol de gran importancia en la unión al sustrato por parte de las enzimas ARNasa III, asegurando la eficiencia de la reacción de corte. Como describimos anteriormente, DCL1 presenta dos dsRBDs en tándem en el extremo C-terminal, DCL1-A y B. Estos dominios son esenciales para el funcionamiento de la proteína *in vivo*. En plantas mutantes que generan transcritos truncados de DCL1 en el segundo dsRBD (*dcl1-9*), el procesamiento de miARNs se ve fuertemente afectado. Estas plantas tienen un fenotipo marcado que se caracteriza por el arresto de la región central del meristema floral en un estado indeterminado, óvulos defectuosos y la morfología de la mayoría de los órganos se ve afectada (Figura 1.14) (Jacobsen et al. 1999). Por otro lado, una mutante insercional en la cual

se ha interrumpido la secuencia de DCL1 en la región del primer dsRBD (mutante *dcl1-6*) es letal a nivel del embrión, probablemente por abolir la función de la enzima (Schauer et al. 2002). Esto sugiere que al menos un dominio dsRBD es indispensable para que la enzima sea capaz de realizar su función *in vivo*.

El fenotipo aberrante de las plantas mutantes *dcl1-9* se correlaciona con un procesamiento erróneo de los precursores de miARN, lo cual resulta en una liberación de miARNs con secuencias anómalas que no son capaces de llevar a cabo la función de los silvestres. La enzima con un solo dominio dsRBD C-terminal es capaz de procesar precursores, efectuando cortes sucesivos separados por 21 nt entre sí. Esto quiere decir que la actividad Dicer de DCL1 no se ve alterada por la ausencia del segundo dsRBD. El procesamiento erróneo se origina en que el sitio del primer corte no es reconocido correctamente por esta mutante, por lo tanto el registro que define el miARN maduro liberado al cabo de los cortes secuenciales se ve desfasado respecto del procesamiento silvestre (Kurihara & Watanabe 2004). Es importante destacar que estas conclusiones se basan en el análisis de un solo precursor, el pri-miR163.

Haciendo uso de esta misma línea mutante, Watanabe y sus colaboradores propusieron que DCL1 interacciona con la proteína accesoria HYL1 a través de DCL1-B. La mutante *dcl1-9* es incapaz de co-inmunoprecipitar con HYL1, mientras que la proteína silvestre sí puede hacerlo (Kurihara et al. 2006). En este mismo trabajo, los autores observaron que las plantas deficientes en HYL1 (*hyl1-2*) presentan procesamiento aberrante de precursores de miARN. A pesar de esto, son capaces de acumular miARNs maduros con menor eficiencia que las plantas silvestres, pero significativamente mayor que las plantas *dcl1-9*. Esto indica que la presencia de HYL1 es parcialmente necesaria para el procesamiento correcto de los precursores, indicando que la interacción con DCL1-B no es indispensable para el funcionamiento de DCL1.

En un estudio sobre la localización intracelular de las proteínas de procesamiento, Fang y Spector demostraron que DCL1 y HYL1 interactúan entre sí y co-localizan en cuerpos específicos dentro del núcleo, los cuerpos D (Fang & Spector 2007). Además, los autores demostraron que la mutante *dcl1-9* localiza de forma difusa en el nucleoplasma y no interacciona con HYL1 en los cuerpos D, sugiriendo nuevamente la interacción entre las dos proteínas a través del dsRBD C-terminal de DCL1. La interacción entre las proteínas fue ensayada en células vivas empleando la técnica de complementación bimolecular de la fluorescencia (BiFC). En este trabajo se demuestra que la presencia de DCL1-B no es necesaria para la entrada de la enzima al núcleo, pero sí para su localización en los cuerpos D.

En un estudio más reciente, se probó que las mutantes *dcl1-9* no pueden ser complementadas por la expresión de los dominios dsRBD por separado, demostrando que esta región es una parte integral de la enzima (Liu et al. 2013). De este trabajo se desprende también que, en mutantes *hyl1-2*, DCL1 sigue formando parte de los cuerpos D, por lo tanto la interacción entre estas dos proteínas no es esencial para la localización de DCL1. Un dato relevante de este trabajo consiste en que tanto una construcción expresando ambos dominios dsRBD de DCL1 (DCL1-AB) como DCL1-B aislado localizan en los cuerpos D, mientras que DCL1-A aislado no lo hace. La observación principal de este trabajo consiste en que la construcción DCL1-AB es capaz de complementar el fenotipo de plantas *hyl1-2*, por interacción de DCL1-B de la enzima con DCL1-B del doble dominio libre. Los autores plantean que en el arreglo de dos dominios dsRBDs en tándem, tanto de DCL1 como de HYL1, uno de los dominios está especializado en la unión a ARNdh mientras que el otro se encarga de la interacción con otros dominios proteicos. Según este modelo, el reconocimiento correcto de los precursores de miARN por parte de DCL1 requeriría la presencia de este arreglo por duplicado, uno como parte integral de la enzima y



otro externo. El arreglo externo silvestre correspondería a HYL1 interaccionando con DCL1-B, que puede ser reemplazado por el doble dominio de DCL1 expresado independientemente.

### 1.7. Reconocimiento del sitio de corte sobre precursores de miARNs

En base a la evidencia disponible hasta la fecha, no disponemos de información suficiente que nos brinde una clara descripción de la forma en que DCL1 reconoce específicamente el sitio donde realizar el primer corte sobre los precursores de miARN. Este sitio es el de mayor importancia en la biogénesis de los miARN ya que es el que define el registro a partir del cual se realizan los cortes sucesivos posteriores a una distancia fija de 21 nt entre sí. Habiendo revisado el mecanismo de reconocimiento del sustrato por parte de otras enzimas ARNasa III podemos hipotetizar que el evento del primer corte por parte de DCL1 no involucra al dominio PAZ, ya que no se dispone de un extremo 3' protuberante en la estructura de los pri-miARNs. Respecto a esto destacamos que ni Drosha ni las ARNasas III bacterianas cuentan con dominios PAZ. Por otro lado, si consideramos que la orientación del sustrato dentro de la enzima es probablemente análoga a la que se propone para Dicer de humanos (Figura 1.12), el dominio helicasa estaría ubicado hacia el interior del tallo en el contexto de un precursor largo como es el caso en los precursores procesados por DCL1. Por lo tanto, probablemente el dominio helicasa tampoco estaría involucrado en el reconocimiento específico determinante del sitio del primer corte sobre el precursor. En este sentido resulta interesante que una mutante de DCL1 denominada *dcl1-7*, que presenta un dominio helicasa inactivo, no es capaz de acumular miARNs maduros pero presenta una alta eficiencia y precisión en la reacción del primer corte (Kurihara et al. 2006).

La enzima ARNasa III de *E. coli* produce fragmentos de 12 pb a partir de ARN doble hebra (Ma et al. 2012), Drosha corta a una distancia fija de 11 pb (Han et al. 2006), mientras una construcción que incluye los dominios ARNasa III y el dsRBD de Dicer de humanos produce fragmentos de 15 pb a partir del sustrato ARNdH (Ma et al. 2012). Sintetizando esta información con la definición del determinante de 15 pb sobre la estructura calculada de los precursores del grupo del Dr. Palatnik (Bologna et al. 2013), podemos hipotetizar que el sitio del primer corte sobre los pri-miARN depende específicamente de la forma en que los dominios ARNasa III junto con los dsRBDs reconocen al precursor en la formación inicial del complejo con el sustrato y las proteínas accesorias, de forma independiente del dominio PAZ y el helicasa. Por esto llegamos a considerar un posible rol crucial para los dsRBDs C-terminal de DCL1 en la formación del complejo inicial y definición del sitio del primer corte. En conjunto, estas observaciones nos motivaron a abordar el problema desde dos aristas diferentes, una caracterización exhaustiva de los dominios de interacción con el ARNdH sustrato de DCL1 y la determinación experimental de la estructura de los precursores de miARN en plantas. Los resultados obtenidos no sólo aportarían información respecto al mecanismo de silenciamiento génico en plantas, sino que también podrían aportar información valiosa respecto a la biofísica del reconocimiento molecular por parte de los dominios de unión a ARNdH, así como también una complementación de los estudios basados en estructuras calculadas de los pri-miARNs.

## 1.8. Objetivos

Habiendo revisado las evidencias disponibles sobre la función de los dominios dsRBD C-terminal de DCL1, es claro que hasta el momento no disponemos de una descripción estructural detallada que expliquen la función exacta de estos dominios en el contexto del complejo de procesamiento. Podemos concluir que esta región es indispensable para la función de DCL1 *in vivo*. Respecto a DCL1-B podemos decir que su función está probablemente especializada en la localización de DCL1 dentro del núcleo y la interacción con otros componentes del complejo de procesamiento. En cuanto a DCL1-A se dispone de menos información. Sabemos que no está implicado en la localización de la enzima dentro del núcleo y que en ausencia de DCL1-B no es suficiente para posicionar a la enzima sobre el precursor en el sitio correcto, aunque sí es suficiente para posibilitar la actividad Dicer de DCL1. Su rol probablemente sea de importancia para la interacción con el sustrato y la eficiencia de la catálisis.

Como **objetivo general** de este trabajo nos propusimos estudiar el reconocimiento de precursores de miARNs por la enzima DCL1.

Para este fin planteamos los siguientes **objetivos particulares**:

- Optimización de la expresión y purificación de DCL1-A para realizar la caracterización biofísica de la proteína, principalmente por RMN.
- Descripción del ensamble conformacional de la forma libre de DCL1-A.
- Descripción de la forma de DCL1-A en complejo con el sustrato y su mecanismo de unión y reconocimiento.
- Validación experimental de determinantes estructurales del procesamiento de precursores de miARNs.

## 2. CAPITULO 2: Materiales y Métodos

### 2.1. Cepas, plásmidos y medios de cultivo

#### 2.1.1. Cepas bacterianas

*Escherichia coli* BL21 (DE3): genotipo  $F^- ompT gal dcm lon hsdS B (r_B^- m_B^-) \lambda(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])$ . El profago  $\lambda$  (DE3) incluye el gen que codifica para la ARN polimerasa del fago T7 bajo control del promotor *lacUV5*, y el gen *lacI*, que codifica para el represor de la transcripción LacI. Utilizamos esta cepa para la expresión de proteínas recombinantes.

*Escherichia coli* DH5 $\alpha$  (Hanahan 1983): usamos esta cepa para guardar los plásmidos obtenidos en un cepario. Su genotipo es: *fhuA2  $\Delta$  (argF-lacZ) U169 phoA glnV44  $\Phi$ 80  $\Delta$ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*.

#### 2.1.2. Genes sintéticos

Adquirimos genes sintéticos a través de la empresa GeneScript (<http://www.genscript.com/>). Suministramos la secuencia de aminoácidos a expresar. La empresa se encargó de la optimización de la secuencia de ADN para la expresión en *E. coli*, la síntesis del fragmento de ADN correspondiente incluyendo sitios de restricción para las enzimas NdeI y BamHI en los extremos, y su clonado en el vector pUC-57. Las preparaciones plasmídicas se entregan en forma liofilizada. Previamente a la transformación de bacterias competentes resuspendimos las muestras en 100  $\mu$ l de agua miliQ estéril.

#### 2.1.3. Vectores plasmídicos

Plásmido pET-TEV (Kapust et al. 2001): lo utilizamos como vector de expresión de proteínas recombinantes. Es un plásmido derivado del pET28-a al cual se le eliminó el sitio de corte de trombina y se reemplazó por el sitio de corte de la proteasa viral TEV. Presenta un sitio de iniciación de la transcripción para la T7 ARN Polimerasa, resistencia a kanamicina como marcador de selección y una secuencia codificante para una etiqueta de histidina.

Plásmido pUC-57: Los genes sintéticos sintetizados por la empresa GeneScript son suministrados dentro del vector pUC-57. Este vector contiene un gen de resistencia al antibiótico ampicilina como marcador de selección y el fragmento N-terminal del gen de la  $\beta$ -galactosidasa de *E. coli* (*lacZ*) interrumpido por el sitio de múltiple clonado, lo que permite seleccionar las colonias que incorporan insertos en cepas que contienen la región C-terminal de la  $\beta$ -galactosidasa.

#### 2.1.4. Medios de cultivo

Para las purificaciones de proteínas utilizamos dos medios de cultivo diferentes en función del tipo de experimento a realizar:

Expresión de proteína marcada: Medio mínimo M9 (en adelante, M9):  $\text{KH}_2\text{PO}_4$  25 mM,  $\text{Na}_2\text{HPO}_4$  50 mM, NaCl 10 mM,  $\text{MgCl}_2$  1 mM,  $\text{CaCl}_2$  0,2 mM, Glucosa 0,4% p/v, Solución de sales 1 ml/l ( $\text{HCl}$  concentrado 51,3 ml/l,  $\text{MgCl}_2$  10,75 g/l,  $\text{CaCO}_3$  2 g/l,  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  4,5 g/l,  $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$  1,12 g/l,  $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$  280 mg/l,  $\text{H}_3\text{BO}_3 \cdot \text{H}_2\text{O}$  60 mg/l,  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  1,44 g/l) y 1 g/l  $\text{NH}_4\text{Cl}$ . Para la obtención de muestras con marcación isotópica utilizamos 1 g/l de  $^{15}\text{NH}_4\text{Cl}$  o 1 g/l de  $^{15}\text{NH}_4\text{Cl}$  y

2 g/l de [U-<sup>13</sup>C]-glucosa (Cambridge Isotope Laboratories). Para la obtención de muestras perdeuteradas preparamos el medio M9 en 100% agua deuterada (<sup>2</sup>H<sub>2</sub>O) y se utilizamos 1 g/l de <sup>15</sup>NH<sub>4</sub>Cl o 1 g/l de <sup>15</sup>NH<sub>4</sub>Cl y 2 g/l de [U-<sup>2</sup>H-<sup>13</sup>C]-glucosa.

## 2.2. Transformación bacteriana

### 2.2.1. Producción de células competentes

Preparamos bacterias competentes de las cepas de *E. coli* descritas en el apartado 2.1.1. Para esto, crecimos las células en 500 ml de LB hasta alcanzar una DO<sub>600</sub> de ca. 0,5. Enfriamos las células en hielo y centrifugamos a 2600 g durante 10 min a 4°C. Resuspendimos el sedimento celular en 100 ml de una solución fría de MgSO<sub>4</sub> 100 mM estéril e incubamos por 30 minutos en hielo. Posteriormente, recuperamos las células por centrifugación a 2600 g, a 4°C durante 10 min y resuspendimos el sedimento obtenido en 10 ml de una solución fría y estéril de CaCl<sub>2</sub> 100 mM, glicerol 15% v/v. Por último, fraccionamos el volumen total en muestras de 300 µl que congelamos y conservamos a -80°C hasta el momento de la transformación.

### 2.2.2. Transformación de *E. coli*

Realizamos la transformación utilizando bacterias competentes generadas como describimos en el apartado anterior. Tomamos alícuotas de 100 µl de bacterias competentes; adicionamos 1 µl de plásmido o 5 µl de mezcla de ligación según correspondiera. En el caso de los controles negativos agregamos 1 o 5 µl de agua como reemplazo del plásmido. Incubamos en hielo durante una hora, luego 90 segundos a 42°C y finalmente 2 min en hielo. A continuación adicionamos 900 µl de medio LB e incubamos durante una hora a 37°C. Transcurrida la incubación centrifugamos el cultivo celular 2 min a 2600 g a temperatura ambiente, resuspendimos el sedimento de células en 100 µl de medio LB y finalmente sembramos en placas de medio LB-agar preparadas con el antibiótico correspondiente. Incubamos las placas toda la noche a 37°C.

### 2.2.3. Selección y análisis de transformantes

Realizamos la selección de las colonias *E. coli* transformadas suplementando las placas con medio LB con el antibiótico adecuado según el vector usado. Cuando el vector fue pUC-57, suplementamos con 100 µg/ml de ampicilina y para el vector pET-TEV con 50 µg/ml de kanamicina.

## 2.3. Subclonado de fragmentos de ADN

### 2.3.1. Minipreparaciones de ADN plasmídico

Purificamos plásmidos utilizando el “kit” de purificación de ADN Wizard® SV Gel (Promega), siguiendo las indicaciones del fabricante.

### 2.3.2. Estimación de la concentración de ADN

Estimamos la concentración de ácidos nucleicos adquiriendo espectros de absorción entre 340 y 240 nm en un espectrofotómetro Jasco V-530 con cubetas de paso óptico de 1 cm. Analizamos los espectros con el programa Jasco Spectra Manager. Calculamos la concentración considerando que 1 unidad de absorbancia a 260 nm de ADN corresponde a una concentración de 50 µg/ml.

### 2.3.3. Digestión de ADN con endonucleasas de restricción

Realizamos las digestiones de ADN con endonucleasas de restricción utilizando enzimas y tampones comerciales. Las condiciones de reacción fueron las especificadas por el fabricante.

### 2.3.4. Electroforesis de ADN en geles de agarosa

Realizamos la separación de moléculas de ADN por electroforesis en geles de agarosa mediante el sistema tipo submarino (Fritsch et al. 1989) en cubas de acrílico. Utilizamos concentraciones de agarosa de 1 a 2,5% p/v según el tamaño de las moléculas de ADN a resolver. Preparamos los geles con solución amortiguadora TAE 1X (Tris-HCl 40 mM, Ácido acético glacial 20 mM y EDTA 1 mM) y agregamos el colorante de ADN SYBR® Safe (Invitrogen). Previo a la siembra, mezclamos las muestras de ADN con 1/6 volúmenes de solución de siembra (glicerol 30%, Xilene cyanol 0,25% p/v, az  $\mu$ l de bromo fenol 0,25% p/v). Llevamos a cabo las corridas electroforéticas en solución amortiguadora TAE 1X, empleando voltaje constante. Utilizamos marcadores de tamaño molecular comerciales adecuados a la longitud de los fragmentos a visualizar.

### 2.3.5. Purificación de fragmentos de ADN a partir de geles de agarosa

Luego de la separación electroforética, identificamos el fragmento de ADN deseado mediante la visualización del gel con un transiluminador UV y escindimos la banda correspondiente con bisturí. Llevamos a cabo la purificación de los fragmentos de ADN del gel utilizando el “kit” de purificación AxyPrep DNA Gel Extraction Kit (Axygen).

### 2.3.6. Ligación de fragmentos de ADN

Llevamos a cabo la ligación de fragmentos de ADN generados por corte con enzimas de restricción en los vectores plasmídicos correspondientes con la enzima T4 ADN ligasa (Fermentas). Las condiciones de la reacción fueron las especificadas por el fabricante. En todos los casos el volumen de reacción de ligación fue de 10  $\mu$ l y utilizamos 5  $\mu$ l de la misma para transformar *E. coli* DH5 $\alpha$ .

#### 2.3.6.1. Control de transformantes: análisis por mapa de restricción

Para confirmar la presencia de inserto del tamaño esperado, realizamos el corte con BamHI y NdeI, de acuerdo al protocolo descrito por el proveedor de la enzima. Analizamos los resultados mediante electroforesis en geles de agarosa 1% p/v. Estimamos los tamaños de los fragmentos de ADN utilizando el marcador de peso molecular 100 pb Ladder (Promega).

#### 2.3.6.2. Secuenciación de ADN

Obtuvimos la secuencia de los insertos a través del servicio de secuenciación de la Universidad de Maine (USA). Una vez obtenida la secuencia, verificamos la identidad de la misma utilizando el programa de acceso libre Ape (A Plasmid Editor) (<http://biologylabs.utah.edu/jorgensen/wayned/ape/>).

## 2.4. Expresión y purificación de proteínas

### 2.4.1. Expresión de proteínas recombinantes

Los plásmidos de expresión para DCL1 y HYL1 fueron construidos por el Dr. Nicolás Bologna en el marco de una colaboración con el grupo del Dr. Javier Palatnik. En todos los casos, los

dominios fueron definidos a partir de la herramienta PROSITE (<http://prosite.expasy.org/>). La construcción que codifica para el dsRBD1 de DCL1, DCL1-A, incluye los residuos M1732-N1811 según la secuencia de la enzima DCL1 completa (UniProt ID: Q9SP32). La construcción DCL1-B, que codifica para el segundo dsRBD de DCL1, incluye los residuos N1826-S1909. El doble dominio DCL1-AB incluye los residuos H1733-S1909. En este trabajo produjimos además las construcciones DCL1-AN (G1707-G1810) y DCL1-ARb (D1709-N1811). En el caso del dsRBD1 de HYL1 (HYL1-A), la construcción incluye los residuos M1-K103 de la secuencia de HYL1.

A partir de genes sintéticos subclonamos en el vector pET-TEV las construcciones correspondientes a los dsRBD1 de DCL3 y DCL4 de *A. thaliana*, y el de Dicer de ratón. La construcción correspondiente al dominio de DCL3, DCL3-A, consiste en los residuos L1365-S1447 (UniProt ID: Q9LXW7), y en el caso de DCL4, DCL4-A, corresponde a la región I1461-K1543 (UniProt ID: P84634). Para la obtención del dsRBD de Dicer de ratón utilizamos la secuencia correspondiente a los residuos R1845-Q1910 de la secuencia anotada de la proteína completa (UniProt ID: Q8R418).

Para la expresión de las proteínas en general partimos de una placa fresca de transformación de células de *E. coli* BL21 (DE3), tomamos una colonia aislada, inoculamos 2 ml de medio LB e incubamos durante 6 horas a 37°C con agitación constante a 220 rpm. A partir de este pre-cultivo inoculamos 500 µl en 50 ml de medio M9 (para proteínas con marca isotópica) o LB (para proteínas sin marca) e incubamos por 12 horas a 37°C con agitación constante. Finalmente, trasvasamos los 50 ml de este cultivo a 1 l de medio M9 o LB, según correspondiera. Incubamos estos cultivos a 37°C y en agitación hasta alcanzar una OD<sub>600</sub> de ca. 0,6. En este punto, indujimos la expresión de las proteínas recombinantes con IPTG 1 mM y continuamos la incubación del cultivo durante 4 horas a 37°C con agitación.

#### 2.4.1.1. Purificación de proteínas en condiciones nativas

Llevamos a cabo la purificación las construcciones DCL1-A, DCL1-B, DCL1-AB, DCL1-AN, y el dsRBD de Dicer de ratón en condiciones nativas utilizando las siguientes soluciones:

A: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 5 mM y β-mercaptoetanol 1 mM

B: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 25 mM y β-mercaptoetanol 1 mM

C: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 350 mM y β-mercaptoetanol 1 mM

D: Tris 50 mM (pH 8), NaCl 500 mM y β-mercaptoetanol 1 mM

E: Na<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub> 100 mM pH 7, NaCl 50 mM, β-mercaptoetanol 5 mM

F: Tris 50 mM (pH 8), NaCl 50 mM y β-mercaptoetanol 1 mM

Cosechamos las células inducidas mediante centrifugación a 6000 g por 15 min. Resuspendimos el sedimento celular en 20 ml de la solución A y lisamos las células utilizando un sonicador. Aplicamos 6 pulsos de 30 segundos con un intervalo de 1 minuto entre cada pulso. Con el objetivo de recuperar las proteínas en solución, centrifugamos el extracto de lisis a 20.000 rpm en rotor ss-34 (Sorvall) durante 1 hora a 4°C. Sembramos el sobrenadante obtenido en una columna que contenía 2 ml de resina de “Sepharosa Ni-NTA superflow” (QUIAGEN), previamente equilibrada con 5 volúmenes de resina de solución A. Lavamos la resina con 20 ml de solución B para eliminar las proteínas unidas inespecíficamente. Por último eluimos la proteína de interés con solución C. Juntamos las fracciones eluidas que contenían proteína y las dializamos contra 1 l de solución D para eliminar el imidazol. Previamente agregamos a la

muestra proteasa TEV con etiqueta de Histidina para que la digestión ocurra simultáneamente con la diálisis durante toda la noche a 4°C. A continuación separamos la proteasa de la proteína de interés sembrando nuevamente en una columna de resina de Ni(II) equilibrada en solución A. Para recuperar la mayor parte de la proteína sin etiqueta de histidina lavamos la resina con 20 ml de solución A y luego con 15 ml de solución B. Juntamos los lavados y concentramos empleando concentradores de membrana de corte por peso molecular de tamaño apropiado para cada construcción (3 KDa o 10 KDa, Amicon) centrifugando a 6000 g a 4°C. En el caso de la construcción DCL1-A realizamos un paso extra de purificación que consistió en sembrar la muestra en una columna de exclusión molecular “Superdex G75” (GE) equilibrada en solución E. Colectamos las fracciones conteniendo la proteína y concentramos. En el caso de las construcciones DCL1-B y DCL1-AB necesitamos separarlas de ácidos nucleicos que copurifican. Para esto dializamos contra solución E, concentramos y separamos por una cromatografía de intercambio catiónico con una columna de “Sephadex CM” (GE). Equilibramos la columna con solución E y se eluimos utilizando un gradiente de 50 mM a 1 M NaCl.

En todos los casos al finalizar la purificación pasamos las muestras a la solución reguladora de pH necesaria para los experimentos siguientes empleando columnas de desalado PD-10 (GE) según las instrucciones del fabricante.

#### 2.4.1.2. Purificación de proteínas en condiciones desnaturalizantes

Para las proteínas HYL1-A, DCL3-A, DCL4-A y DCL1-ARb realizamos purificaciones en condiciones desnaturalizantes para lo cual utilizamos las siguientes soluciones:

I:  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  100mM, Tris-HCl 10 mM, urea 8 M, pH 8,0

II:  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  100mM, Tris-HCl 10 mM, urea 8 M, pH 6,3

III:  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  100mM, Tris-HCl 10mM, urea 8 M, pH 4,5

La cosecha de células y la lisis celular las realizamos en las mismas condiciones que en el apartado anterior, excepto que en este caso, resuspendimos el sedimento celular en 20 ml de la solución I. Sembramos el sobrenadante obtenido luego de la centrifugación a 20.000 rpm en rotor ss-34 (Sorvall) utilizando una columna de afinidad de Ni(II) previamente equilibrada con 5 volúmenes de columna de solución I. Con el objetivo de eliminar proteínas unidas de manera inespecífica, lavamos la columna con 20 ml de solución II. Por último eluimos la proteína de interés con solución III, en fracciones de 1 ml. Luego procedimos al replegado de las proteínas obtenidas como se describe a continuación.

#### 2.4.2. Replegado de proteínas

Para el replegado dializamos las fracciones que contenían proteína contra 100 volúmenes de solución de replegado ( $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  100 mM (pH=7.0), NaCl 50 mM,  $\beta$ - mercaptoetanol 1 mM, Arg 50 mM, Glu 50 mM) por 12 horas. En todos los pasos de diálisis utilizamos membranas de diálisis de corte 10 KDa pre-tratadas por ebullición durante una hora en agua destilada y EDTA para remover las trazas de metales pesados. Luego de dializar las proteínas las concentramos utilizando concentradores Amicon como se describió previamente. Finalmente en los casos que fue necesario cambiamos la solución reguladora de pH según los requerimientos de los experimentos siguientes empleando columnas PD-10 (GE).

#### 2.4.3. Purificación de la proteasa TEV

Para la purificación de la proteasa TEV utilizamos las siguientes soluciones:

O: Tris 50 mM (pH 8), NaCl 500 mM, Glicerol 10 %, Imidazol 20 mM y  $\beta$ -mercaptoetanol 1 mM

P: Tris 50 mM (pH 8), NaCl 1,5 M, Glicerol 10 %, Imidazol 100 mM y  $\beta$ -mercaptoetanol 1 mM

Q: Tris 50 mM (pH 8), NaCl 200 mM, Glicerol 10 %, EDTA 2 mM y  $\beta$ -mercaptoetanol 1 mM.

Cosechamos las células inducidas mediante centrifugación a 6000 g por 10 minutos. Resuspendimos el sedimento celular en 25 ml de la solución O y a continuación sonicamos de la misma forma que en el apartado 2.4.1. Sembramos el sobrenadante en una resina de afinidad de Ni (II) y realizamos un lavado con la solución P. Eluimos con un gradiente de Imidazol de 100 a 500 mM. A continuación, juntamos las fracciones que contenían proteína y agregamos DTT y EDTA 1mM cada uno y dializamos contra la solución Q durante toda la noche a 4°C. Finalmente, agregamos Glicerol 10 % y DTT 10 mM, fraccionamos y almacenamos en un freezer -20°C.

Evaluamos la actividad de la proteasa utilizando la proteína HYL1-A fusionada a una etiqueta de histidina con sitio de corte de TEV. Preparamos mezclas de reacción con distinta relación proteasa:proteína de fusión, desde 1:300 hasta 1:12. Dejamos transcurrir la digestión a 4°C durante toda la noche y evaluamos la actividad en geles de poliacrilamida 15%.

#### 2.4.4. Estimación de la concentración de proteínas

Estimamos la concentración de proteína adquiriendo espectros de absorción entre 320 y 240 nm en un espectrofotómetro Jasco v-530 con cubeta de paso óptico 1 cm. Analizamos los espectros en el programa Jasco Spectra Manager. En cada caso, empleamos el coeficiente de extinción molar a 280 nm obtenido con el servidor "Protparam" (<http://www.expasy.org/tools/protparam.html>), utilizando la secuencia primaria de cada proteína.

#### 2.4.5. Electroforesis de proteínas en geles de poliacrilamida

Llevamos a cabo los controles de todas las etapas de purificación mediante electroforesis en geles de poliacrilamida en condiciones desnaturalizantes y reductoras. El sistema discontinuo consiste en un gel de concentración (poliacrilamida 4,5% en Tris-HCl 0,126 M, pH 6,8, SDS 0,26% p/v) seguido por un gel de separación (poliacrilamida 15% en Tris-HCl 0,36 M, pH 8,8, SDS 0,26% p/v). Mezclamos las muestras con solución de siembra (glicerol 5%, SDS 2%,  $\beta$ -mercaptoetanol 0,1%, az  $\mu$ l de bromofenol 0,1 mg/ml) y posteriormente las calentamos durante 5 minutos a 100°C. Para la corrida empleamos una solución amortiguadora de Tris 0,3% p/v, glicina 1,44% p/v y SDS 0,1% p/v y aplicamos un potencial constante de 180 V. En todos los casos utilizamos el sistema Miniprotean 4 (Bio-Rad Laboratories). Luego de la corrida, fijamos las proteínas tratando el gel durante 5 minutos en una solución decolorante [etanol: ácido acético: agua (30:10:60)] con agitación y teñimos con una solución de Az  $\mu$ l Brillante de Coomásie R250 1 % p/v en etanol: ácido acético: agua (50:10:40). Finalmente decoloramos el gel para visualizar las bandas manteniéndolo en solución decolorante con agitación constante.

### 2.5. Espectroscopia de resonancia magnética nuclear

#### 2.5.1. Fundamentos teóricos

##### 2.5.1.1. El espectro $^1\text{H}^{15}\text{N}$ -HSQC

El experimento  $^1\text{H}^{15}\text{N}$ -HSQC (Heteronuclear Single Quantum Coherence) es la base de gran parte de los estudios de RMN multidimensional en proteínas. En este experimento se

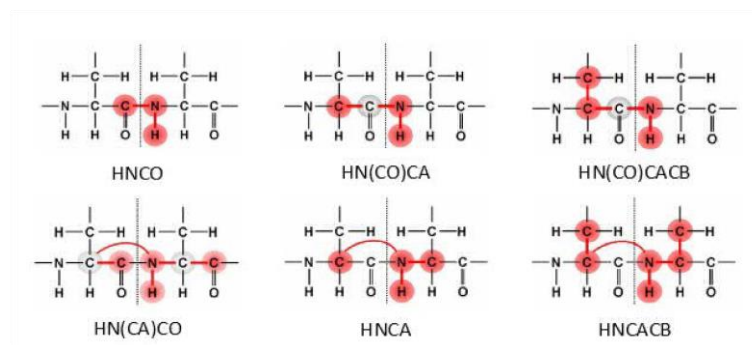


correlaciona el desplazamiento químico de un núcleo de  $^1\text{H}$  con el de un núcleo de  $^{15}\text{N}$  al que se encuentra directamente unido, con una constante de acoplamiento escalar ( $^1J$ ) de 90–94 Hz. Así, en el espectro  $^1\text{H}^{15}\text{N}$ -HSQC de una proteína se espera observar un número total de señales igual al número de residuos de la proteína, sin contar los residuos de prolina, que no poseen  $^1\text{H}$  amídico y, por ende, no generan señal en este espectro, y el residuo amino terminal, que intercambia los  $^1\text{H}$  muy rápido con el solvente. A las señales correspondientes a los grupos amida del esqueleto de la proteína se suman las señales correspondientes a los grupos amida de las cadenas laterales de Gln y Asn y las señales de los pares NH de los anillos indólicos de los residuos de Trp.

Teniendo en cuenta la información que aporta el espectro  $^1\text{H}^{15}\text{N}$ -HSQC, el potencial de dicho experimento para los estudios biofísicos de proteínas depende de que se disponga de la asignación de sus señales. Es decir, para poder interpretar los cambios observados en los espectros adquiridos en diferentes estados de la proteína, es necesario determinar a qué residuo corresponde cada señal.

La posición de cada señal en el espectro depende fuertemente del microentorno químico del par  $^1\text{H}$  -  $^{15}\text{N}$  amida correspondiente. Esta dependencia convierte al espectro  $^1\text{H}^{15}\text{N}$ -HSQC en la huella dactilar de una proteína bajo las condiciones de estudio. Por este motivo, el  $^1\text{H}^{15}\text{N}$ -HSQC es el experimento ideal, en la gran mayoría de los casos, para estudiar la interacción de una proteína con ligandos, así como también el efecto del pH, la temperatura o la fuerza iónica, entre muchas otras aplicaciones. Un experimento equivalente al  $^1\text{H}^{15}\text{N}$ -HSQC que permite correlacionar los desplazamientos químicos de  $^1\text{H}$  con los de un heteronúcleo es el  $^1\text{H}^{15}\text{N}$ -HMQC (Heteronuclear Multiple Quantum Coherence). Este experimento brinda una mayor relación señal:ruido que el  $^1\text{H}^{15}\text{N}$ -HSQC y puede ser utilizado en los casos en que la resolución no sea un punto crítico. Para moléculas de tamaño molecular relativamente grande se emplea la variación  $^1\text{H}^{15}\text{N}$ -HSQC-TROSY. En este caso se obtiene la misma información pero la secuencia de pulsos está diseñada de forma de seleccionar una componente de la señal proveniente de los núcleos acoplados que presenta mayor intensidad y menor ancho de línea, lo cual incrementa la resolución obtenida.

#### 2.5.1.2. Asignación de resonancias $^1\text{H}$ , $^{15}\text{N}$ y $^{13}\text{C}$



**Figura 2.1** Esquemas de los experimentos de triple resonancia para asignar el esqueleto peptídico

*En rojo se indican los núcleos que son detectados y en gris se indican aquellos a través de los cuales se filtra la magnetización. Las líneas rojas indican los caminos de transferencia de la magnetización.*

Para llevar a cabo la asignación de las señales de un  $^1\text{H}^{15}\text{N}$ -HSQC, la estrategia más utilizada consiste en la adquisición de experimentos de triple resonancia. En estos experimentos, cada par  $^1\text{H}$  -  $^{15}\text{N}$  es correlacionado con el valor de desplazamiento químico de distintos núcleos de

$^{13}\text{C}$  del mismo residuo y/o de su residuo anterior en la secuencia proteica, según el caso. El éxito de esta estrategia se basa en que los desplazamientos químicos de  $^{13}\text{C}$  de los carbonos  $\text{C}\alpha$  y  $\text{C}\beta$  de un dado aminoácido son más dependientes de la identidad química del mismo que de su entorno, posibilitando la identificación del tipo de residuo. Algunos de los experimentos utilizados para la asignación de las resonancias del esqueleto peptídico de proteínas se muestran en la Figura 2.1.

#### 2.5.1.3. Cálculo de estructura empleando CS-Rosetta

Para resolver la estructura tridimensional de una proteína en solución, es necesario asignar las resonancias provenientes de todos los núcleos de la misma, principalmente, aquellos de las cadenas laterales. Además es necesario obtener restricciones de distancias entre los distintos átomos a través del efecto nuclear Overhauser (NOE). Al momento de resolver la estructura de una molécula o complejo de gran tamaño la obtención de estos datos puede llegar a ser muy dificultosa. El origen de las dificultades proviene de que el aumento en el tiempo de correlación rotacional de la molécula en estudio resulta en un rápido decaimiento de la señal y por ende en un ensanchamiento y una a baja disminución de la relación señal:ruido.

Para lograr obtener una estructura de un sistema de peso molecular elevado pueden combinarse varias estrategias. En cuanto a la muestra pueden reemplazarse todos los núcleos  $^1\text{H}$  no intercambiables por núcleos de  $^2\text{H}$ , lo cual reduce la relajación de los núcleos excitados por acoplamiento con los  $^1\text{H}$  circundantes, incrementando la relación señal:ruido y reduciendo el ancho de línea de las señales. Obtener un ancho de línea menor resulta en una mayor resolución, lo cual es relevante al trabajar con sistemas de tamaño molecular elevado por la gran cantidad de señales que se solapan. Por otro lado los experimentos de triple resonancia necesarios para asignar las señales provenientes de los núcleos que conforman el esqueleto peptídico poseen una sensibilidad relativamente alta. Esto hace de la asignación de estas resonancias una meta más factible para sistemas de alto peso molecular, respecto de la asignación completa de todos los núcleos.

Dada la dificultad para obtener los datos y la cantidad de tiempo que requiere, se han hecho grandes esfuerzos para definir el mínimo grupo de parámetros necesario para poder calcular estructuras de proteínas de baja resolución. En este sentido se ha estudiado extensamente la información estructural que se puede obtener a partir de los desplazamientos químicos. Estos valores brindan información sobre factores estructurales tales como la conformación del esqueleto peptídico y de las cadenas laterales, estructura secundaria, fuerza de puentes de hidrógeno y la posición de anillos aromáticos (Case, 1995; Cornilescu, Delaglio, & Bax, n.d.; Shen & Bax, 2007; Wagner, Pardi, & Wüthrich, 1983). Esta información es incorporada en el cálculo de estructuras empleando el programa CS-Rosetta (Shen et al., 2008).

La versión original de Rosetta predice estructuras *de novo* a partir de la secuencia de la proteína (Bradley, Misura, & Baker, 2005). Para cada residuo de la proteína, Rosetta selecciona 200 fragmentos de una base de datos de estructuras cristalográficas que presentan similitud en secuencia de aminoácidos y, por ende, representativos de las conformaciones que cada segmento puede explorar en el plegamiento. Luego utiliza estos fragmentos en un proceso de ensamblado tipo Monte Carlo para encontrar posibles plegamientos de baja energía. Este paso se realiza sobre un modelo en el que sólo se encuentran explícitos los átomos pesados del esqueleto peptídico y un pseudo átomo centroide que representa el volumen de cada cadena lateral. Una vez obtenido el modelo de baja energía se incorporan todos los átomos y se procede al refinamiento. El refinamiento final de la estructura emplea una minimización tipo

Monte Carlo acoplada a un campo de fuerza detallado teniendo en cuenta que considera todos los átomos. Durante este paso se buscan las estructuras de baja energía con un empaquetamiento de cadenas laterales complementario cercano al del modelo inicial óptimo (Rohl, Strauss, Misura, & Baker, 2004).

CS-Rosetta utiliza el mismo principio de funcionamiento que Rosetta pero incorpora algunas modificaciones que incrementan la exactitud de la estructura calculada. Una de las modificaciones se encuentra en la base de datos que utiliza CS-Rosetta de la cual selecciona los fragmentos similares a la secuencia incógnita. Los autores del programa crearon una base de datos de 5.665 estructuras cristalográficas conocidas para las cuales calcularon los desplazamientos químicos teóricos con el programa SPARTA (Shen 2007). Esta base de datos es la que luego se utiliza para seleccionar los fragmentos de estructura conocida, pero en este caso la selección no se realiza simplemente por similitud de secuencia, sino que se le da más peso a la similitud en los desplazamientos químicos secundarios. Es decir, se comparan los desplazamientos químicos secundarios experimentales de la proteína en estudio con los calculados para las estructuras de la base de datos. Esta modificación ya permite incorporar restricciones estructurales experimentales al cálculo, respecto del protocolo original de Rosetta. Una vez generada la base de datos de fragmentos específica para la proteína en estudio CS-Rosetta utiliza el mismo método de ensamblado de los fragmentos que Rosetta para generar modelos de todos los átomos a partir de la secuencia, incluyendo los pasos de la generación de la conformación de baja energía del esqueleto peptídico a baja resolución así como también la etapa final de refinado donde se genera el modelo con todos los átomos. De esta manera se generan entre 10.000 y 20.000 modelos. Una vez obtenidos estos modelos, se seleccionan los 5.000 de menor energía y se utiliza SPARTA para predecir los desplazamientos químicos del esqueleto de cada modelo. Seguidamente se incorpora el desvío entre los desplazamientos químicos calculados y los experimentales al cálculo de la energía total de cada modelo para identificar los modelos de menor energía y se evalúa la convergencia de las estructuras comparando la energía total con el RMSD del C $\alpha$  respecto de la estructura de menor energía obtenida. Finalmente se seleccionan las 10 estructuras de menor energía como las más representativas de los datos experimentales.

#### 2.5.1.4. Intercambio químico

Cuando un núcleo de una muestra puede encontrarse en dos entornos diferentes que resultan en distintos desplazamientos químicos, el espectro de RMN resultante dependerá de la relación entre la velocidad de intercambio entre los entornos y la diferencia de frecuencias de resonancia del núcleo en cada entorno. El fenómeno de intercambio puede darse en un sistema simple en el que ocurren cambios conformacionales, o durante una titulación de dos especies en donde las formas libre y unida de cada especie dan lugar a una frecuencia de resonancia diferente para el mismo núcleo. Debido a las magnitudes de las diferencias de frecuencia que se encuentran entre señales de RMN, se pueden presentar tres casos de régimen de intercambio:

- 1- Cuando el proceso de intercambio ocurre en un régimen más lento que la diferencia de los desplazamientos químicos entre las dos formas, cada núcleo afectado podrá asociarse a dos resonancias en el espectro en cuestión, correspondientes al mismo en dos entornos químicos diferentes.
- 2- Cuando el proceso de intercambio ocurre en un régimen rápido en la escala de tiempo de los desplazamientos químicos no se ve afectada la intensidad de la señal resultante pero sí su

posición en el espectro, la cual resultará ser un promedio de los valores de desplazamiento químico esperados en cada entorno químico, pesado según la población relativa de cada uno.

3- En un régimen de intercambio químico intermedio en la escala de los desplazamientos químicos, se observará la coalescencia de ambas señales dando una señal promedio demasiado ancha como para ser detectada en el espectro de RMN; es decir, que tal intercambio se evidenciará por la ausencia de las resonancias esperadas o por un ensanchamiento de las mismas.

#### 2.5.1.5. Ensamblajes conformacionales de proteínas desestructuradas

Las proteínas desestructuradas presentan una elevada flexibilidad respecto de las proteínas plegadas. Para comprender el comportamiento de estas moléculas es esencial lograr una descripción a resolución atómica que contemple su flexibilidad. Esto plantea un problema esencialmente diferente al abordaje clásico de la biología estructural. La resolución de estructuras clásica apunta a la descripción tridimensional de una única conformación. Esta aproximación tiene poca relevancia para proteínas desestructuradas debido a que una única conformación representa uno de un continuo de subestados significativamente poblados. Desde el descubrimiento de las proteínas intrínsecamente desordenadas se han dedicado grandes esfuerzos al desarrollo de descripciones moleculares que representen la heterogeneidad conformacional de estas proteínas. De las técnicas en solución disponibles, la espectroscopia de RMN se ha establecido como una de las técnicas más poderosas para el estudio de proteínas con elevada flexibilidad. Esta técnica reporta parámetros promedio del ensamble conformacional a resolución atómica.

Para lograr una descripción de la superficie de energía libre explorada por proteínas desestructuradas a partir de datos experimentales se han desarrollado una serie de técnicas analíticas que abordan el problema desde diferentes perspectivas (Jensen et al. 2014). Una de estas técnicas consiste en generar ensambles conformacionales independientemente de los datos experimentales de la forma menos tendenciosa posible, con la intención de explorar la mayor parte del espacio conformacional disponible para una dada secuencia. Una vez obtenido este ensamble de conformaciones se selecciona un grupo de conformaciones que en conjunto reproduzcan los datos experimentales. Un ejemplo de esta aproximación es la combinación de los algoritmos Flexible-Meccano (Bernadó et al. 2005) y ASTEROIDS (Ozenne et al. 2012). Flexible Meccano permite la generación del ensamble de conformaciones al azar que luego son seleccionadas con el algoritmo genético ASTEROIDS para la construcción de un subensamble que representa el estado de la cadena en solución.

#### 2.5.1.6. Acoplamientos dipolares residuales

El valor del acoplamiento dipolar entre un par de núcleos ( $D_{ij}$ ) es directamente proporcional al ángulo ( $\Theta$ ) entre el vector entre los dos núcleos y el vector del campo magnético externo. En solución isotrópica este valor se promedia a cero debido al promediado de todas las orientaciones posibles de la molécula (múltiples ángulos  $\Theta$ ). Si una proteína es disuelta en un medio líquido cristalino orientado diluido, el promediado de las orientaciones deja de ser uniforme porque algunas orientaciones tendrán mayor probabilidad de ocurrir que otras, es decir, las moléculas se encontrarán parcialmente alienadas respecto del campo magnético. Este alineamiento parcial resulta en un valor de  $D$  distinto de cero denominado acoplamiento dipolar residual (RDC, del inglés *residual dipolar coupling*) (Tolman & Ruan 2006).

Los RDCs son sondas del orden angular en una cadena, brindando información de orientaciones relativas a largo alcance de las diferentes regiones de una molécula. En el caso de las proteínas desestructuradas los RDCs brindan información muy valiosa sobre el comportamiento conformacional de estas moléculas en solución. En el caso de los RDCs HN-N medidos para los grupos amida de proteínas desplegadas los valores son cercanos a cero y de signo negativo, con forma de campana invertida debido a la mayor heterogeneidad conformacional de los extremos de la cadena. Las regiones cercanas a prolinas, así como también las regiones que presentan estructuras helicoidales transitorias presentan valores de RDCs positivos, reportando una mayor rigidez respecto del resto de la cadena.

Como los RDCs dependen tanto de la estructura local como global de la proteína, su análisis es un poco más complejo que el de los desplazamientos químicos. Para las proteínas plegadas puede calcularse un tensor de alineamiento, que describe la orientación de la molécula respecto del campo externo, y la orientación de cada vector internuclear relativa al sistema de ejes principal de dicho tensor de alineamiento (Bax & Grishaev 2005). Esta aproximación asume que el tensor de alineamiento puede ser considerado invariable para todas las moléculas de la proteína plegada en la muestra y que el tensor es independiente del comportamiento conformacional de cada vector internuclear. Estas suposiciones son claramente no aplicables a proteínas desestructuradas (Jensen et al. 2014). Empleando ensambles conformacionales como la descripción del estado conformacional de las proteínas desplegadas es posible aplicar el análisis descripto para proteínas plegadas para cada conformero dentro del ensamble y luego obtener como resultado el promedio de los valores.

## 2.5.2. Procedimientos experimentales

En general adquirimos los espectros de resonancia magnética nuclear (RMN) a 298 K. Preparamos las muestras agregando 10% v/v de agua deuterada (Cambridge Isotope Laboratories), que se utiliza como referencia interna de frecuencia para la acumulación de los espectros consecutivos. Utilizamos como referencia para los desplazamientos químicos la señal del agua a 4,77 ppm (pH 6,8; 25°C) relativa a DSS, utilizando las relaciones de frecuencia  $^1\text{H}:\text{X}$  del punto cero de acuerdo a Wüthrich y colaboradores (Markley et al. 1998).

### 2.5.2.1. Asignación de resonancias de DCL1-A

Asignamos los desplazamientos químicos de  $^1\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$  del esqueleto peptídico de DCL1-A libre utilizando espectros de triple resonancia [BEST-HNCA/HN(CO)CA, BEST-HNCACB/HN(CO)CACB, BEST-HN(CA)CO/HNCO y HN(CA)HA] (Lescop et al. 2007) que adquirimos en un espectrómetro Bruker Avance II que opera a 600 MHz, empleando una sonda TXI de triple resonancia ( $^1\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$ ) con detección inversa. La muestra consistió en DCL1-A marcada con  $^{15}\text{N}$  y  $^{13}\text{C}$  700  $\mu\text{M}$  y se encontraba en solución reguladora de pH  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  10 mM pH 6,5,  $\beta$ -mercaptoetanol 1 mM, EDTA 10 mM, 1X cOmplete (inhibidor de proteasa, Roche), y 10% v/v  $\text{D}_2\text{O}$ . Depositamos la asignación en la base de datos BioMagResBank (<http://www.bmrb.wisc.edu/>) bajo la identificación: 19105.

La asignación de las resonancias del esqueleto peptídico de DCL1-A en complejo con el sustrato la obtuvimos a partir de espectros de triple resonancia estándar adquiridos en un espectrómetro Varian (Agilent) que opera a 600,13 MHz equipado con una sonda criogénica. La muestra consistió en DCL1-A perdeuterada 400  $\mu\text{M}$  marcada con  $^{15}\text{N}$  y  $^{13}\text{C}$  y ARN pri-miR172a-ti 500  $\mu\text{M}$  en solución reguladora de pH  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  20 mM pH 7, TCEP 5 mM. La asignación completa fue depositada en la base de datos BioMagResBank bajo la identificación: 19104.

#### 2.5.2.2. Cálculo de la estructura de DCL1-A en complejo con el sustrato

Una vez obtenidos los desplazamientos químicos para los núcleos de la cadena principal de DCL1-A en complejo con el sustrato procedimos al cálculo de la estructura implementando el programa CS-Rosetta según las instrucciones de los autores (Shen et al. 2008).

#### 2.5.2.3. Determinación de los NOEs heteronucleares $^1\text{H}^{15}\text{N}$ de DCL1-A

Medimos los NOEs heteronucleares  $^1\text{H}^{15}\text{N}$  adquiriendo los datos con la secuencia de pulsos estándar de Bruker en el espectrómetro Bruker Avance II 600 mencionado anteriormente. La muestra se encontraba en las mismas condiciones que en el caso de la asignación de la proteína libre.

#### 2.5.2.4. Espectros de intercambio ZZ

Adquirimos espectros de intercambio ZZ-EXSY sobre la misma muestra de proteína perdeuterada en presencia del sustrato empleada para la asignación del esqueleto peptídico. Adquirimos espectros sin mezcla y con 200 ms de tiempo de mezcla, en un espectrómetro Varian (Agilent) DirectDrive que opera a 800 MHz, empleando la secuencia estándar de Varian.

#### 2.5.2.5. Determinación de RDCs de DCL1-A

Para la adquisición de los parámetros de acoplamiento dipolares residuales (RDC), alineamos parcialmente una muestra 400  $\mu\text{M}$  de DCL1-A marcada con  $^{15}\text{N}$  en una fase líquida cristalina formada por C8E5 (Pentaethylene glycol octylether) y octanol. Este medio preparado al 5% p/p (relación de la masa de C8E5 respecto de la del solvente) con una relación molar C8E5/octanol de 0.84 crea una fase nemática en presencia del campo magnético del espectrómetro que induce un alineamiento preferencial de la proteína en la dirección del campo magnético (Rückert & Otting 2000). Medimos los valores de RDC para el par de átomos HN-N del enlace peptídico adquiriendo espectros  $^{15}\text{N}$ -IPAP-HSQC en medio isotrópico y anisotrópico utilizando el espectrómetro Bruker 600 MHz equipado con una consola Avance II y una sonda TXI. En estos espectros se obtienen dos picos para cada señal correspondiente a un par HN-N. La separación entre los picos corresponde al valor de la constante de acoplamiento escalar J entre los dos núcleos en el medio isotrópico y la constante de acoplamiento escalar junto con la constante de acoplamiento dipolar (J+D) en el medio alineado. Midiendo la distancia entre los picos en los espectros adquiridos en los dos medios y calculando luego la diferencia de distancia se obtiene el valor del acoplamiento dipolar residual.

#### 2.5.2.6. Interacción proteína-ácidos nucleicos

Realizamos los experimentos de interacción ARN-proteína empleando un espectrómetro Bruker 600 MHz equipado con una consola Avance II y una sonda TXI. Las condiciones fueron  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  10 mM pH 6,5,  $\beta$ -mercaptoetanol 1 mM, 10% v/v  $\text{D}_2\text{O}$ . Previamente calentamos el ARN a 100°C y enfriamos inmediatamente en un baño de hielo y agua.

Procedimos agregando alícuotas de una muestra de pri-miR172a-ti 400  $\mu\text{M}$  sobre una muestra de proteína marcada con  $^{15}\text{N}$  200  $\mu\text{M}$ . A cada paso, adquirimos un espectro  $^1\text{H}$ - $^{15}\text{N}$ -SOFAS-HMQC (Schanda & Brutscher 2005) y un espectro unidimensional de protones incluyendo la región a campo bajo de los protones imino del ARN.

#### 2.5.2.7. Asignación de los protones imino de pri-miR172a-ti

Para la asignación de los protones imino del pri-miR172a-ti adquirimos espectros de  $^1\text{H}$ - $^{15}\text{N}$ -HSQC en abundancia natural de  $^{15}\text{N}$ , y  $^1\text{H}$ - $^1\text{H}$ -NOESY con un tiempo de mezcla de 150 ms en un espectrómetro Bruker 600 MHz equipado con una consola Avance II y una criosonda TCI. La muestra consistió en pri-miR172a-ti 400  $\mu\text{M}$  en solución reguladora de pH de ácido cacodílico 20 mM pH 6. La asignación se llevó a cabo a través de la caminata por los NOEs imino – imino de las guaninas y los uracilos.

#### 2.5.2.8. Determinación de radio hidrodinámico

Para la determinación del radio hidrodinámico empleamos el método de RMN de gradientes de campo pulsados usando dioxano como patrón interno para independizar las medidas de la viscosidad del medio (Wilkins et al. 1999). A una muestra 200  $\mu\text{M}$  de DCL1-A agregamos 0,5  $\mu\text{l}$  de 1,4-dioxano (Sigma-Aldrich) y 10% v/v  $\text{D}_2\text{O}$ . Utilizamos un espectrómetro Bruker 600 MHz equipado con una consola Avance II y una sonda TXI. Adquirimos espectros  $^1\text{H}$  unidimensionales a distintas potencias de gradiente (0 – 95%) empleando el experimento estándar de Bruker.

### 2.6. Preparación de muestras de ARN

#### 2.6.1. Purificación de oligonucleótidos de ADN

Obtuvimos los oligonucleótidos de ADN a emplear en reacciones de PCR de la empresa Sigma. Una vez en el laboratorio procedimos a su purificación para eliminar productos inespecíficos de la síntesis. Sembramos la totalidad de las muestras en geles desnaturalizantes de poliacrilamida entre 10 y 20% p/v, dependiendo de la longitud del oligonucleótido. Luego de correr la electroforesis visualizamos las bandas por la sombra que producen bajo la luz UV. Cortamos las bandas y las extrajimos por incubación durante toda la noche en solución de extracción (Tris 10mM pH 8, EDTA 1mM, NaCl 300mM). Luego precipitamos el ADN disuelto en la solución de extracción, como se describe en el apartado 2.6.2.

#### 2.6.2. Precipitación de ácidos nucleicos con etanol

A la muestra de ácidos nucleicos a precipitar agregamos 1/10 de volumen de una solución NaCl 3 M y 3 volúmenes de Etanol absoluto frío, incubamos 15 min en hielo seco y centrifugamos 30 min a 10.000 rpm a 4°C. Lavamos el pellet resuspendiendo en 1 ml de Etanol 70% v/v frío y recuperamos volviendo a centrifugar durante 20 min a 10.000 rpm a 4°C. Secamos el pellet en un concentrador por vacío a 37°C durante 5 min y resuspendimos en  $\text{H}_2\text{O}$ . Luego de determinar la concentración como se detalla en el apartado 2.6.7 llevamos a la concentración necesaria en cada caso.

#### 2.6.3. Moldes para transcripción *in vitro*

Para transcripción *in vitro* a escala intermedia del fragmento pri-miR172a-ti se utilizaron oligos de ADN hibridados. Los oligos 1 y 2 de la Tabla 2.1 se mezclaron en proporción equimolar, se calentaron a 95°C durante 5 min y luego se dejaron enfriar a temperatura ambiente durante 30 min. Para la transcripción a mediana escala del precursor completo del miR172a se utilizó como molde plásmido linearizado.

Para las transcripciones a pequeña escala de los precursores completos pri-miR172a, pri-miR171a, pri-miR171b, pri-miR319a y pri-miR394a se construyeron moldes por extensión del solapamiento por PCR. Los oligos largos directo y reverso correspondientes a cada precursor (Tabla 2.1) se mezclaron en cantidades equimolares ( $C_i=0,02 \mu\text{M}$  cada uno) junto con 1 U de

polimerasa de ADN Pfu (Promega), solución reguladora comercial con MgSO<sub>4</sub>, 0,2 mM de cada uno de los dNTPs (dCTP, dATP, dGTP, dTTP). El volumen final fue de 50µl. Esta mezcla se sometió a un primer ciclo de elongación en el cual primero se calentó a 95°C 5 min, luego se bajó la temperatura a la temperatura de hibridación de la región solapada entre los oligos largos y finalmente se llevó la temperatura a 68°C durante 10 min para asegurar la elongación de la región solapada a la longitud final del molde. Seguido a este paso se agregó a cada tubo el oligo 3 de la Tabla 2.1 y el oligo reverso corto específico para cada precursor (también detallados en la Tabla 2.1) en concentración final de 0,2 µM cada uno. Se volvió a agregar 1 U de enzima y se llevó a cabo un programa estándar de 30 ciclos con la temperatura de hibridación correspondiente para cada oligo reverso corto específico. Los resultados de las reacciones de PCR se controlaron sembrando las muestras en geles de agarosa como se describe en el apartado 2.3.4. Una vez corroborado que se obtuvo el fragmento del tamaño deseado se procedió a purificar el molde de la reacción de PCR por extracción fenol / cloroformo / alcohol isoamílico y posterior precipitación con etanol como se detalla en el apartado 2.6.4. Las muestras se resuspendieron finalmente en 20µl de agua.

**Tabla 2.1 Oligonucleótidos de ADN**

N°	Nombre	Secuencia (5'- 3')
1	Pri-miR172a-ti_Fw	TAATACGACTCACTATAGGGCTGCTGTGGCATCTTCGGATGCTGCA TCGGC
2	Pri-miR172a-ti_Rev	GCCGATGCAGCATCCGAAGATGCCACAGCAGCCCTATAGTGAGTC GTATTA
3	T7_Fw	CTGTAATACGACTCACTATA
4	Pri-miR172a_Rev	GGAAAGAATAGTCGTTGATTG
5	Pri-miR172a_long_Fw	CTGTAATACGACTCACTATAGGGAGCCACGGTCGTTGTTGGCTGCT GTGGCATCATCAAGATTCACATCTGTTGATGGACGGTGGTGATTG
6	Pri-miR172a_long_Rev	GGAAAGAATAGTCGTTGATTGCCGATGCAGCATCATCAAGATTCTC ATTTTCATAGAGAACTTTGTGGAGAGTGAATCACCACCGTCCATCA AC
7	Pri-miR171a_Rev	AAAGTCCAAAATAGAGACGAG
8	Pri-miR171a_long_Fw	CTGTAATACGACTCACTATAGGATCCGCACTTTAAGCATGAGAGAG TCCCTTTGATATTGGCCTGGTTCACTCAGATCTTACCTGACCACACA CGTAGATATAC
9	Pri-miR171a_long_Rev	AAAGTCCAAAATAGAGACGAGAGAGTACTGAGATATTGGCGCGGC TCAATCAGATAATCTAGAGAGAATAATGTATATCTACGTGTGTGGT CAGG
10	Pri-miR171b_Rev	TTATAATAACTATCTTTGCCATTTTG
11	Pri-miR171b_long_Fw	CTGTAATACGACTCACTATAGGAGTCGAACAAGAGTGGTTTTATGC AAGGTAACGCGAGATATTAGTGCAGTTCAATCAAATAGTCGTCCTC TTAACTC
12	Pri-miR171b_long_Rev	TTATAATAACTATCTTTGCCATTTTGGTTTACCGCGTGATATTGGC ACGGCTCAATCGAACAACACCGTTCTCCATGAGTTAAGAGGACGAC TATTTGATTG
13	Pri-miR319a_Rev	GGAATACAAAAGAGAGAGGG
14	Pri-miR319a_long_Fw	CTGTAATACGACTCACTATAGGAATATATGTAGAGAGAGCTTCC TTGAGTCCATTACAGGTCGTGATATGATTCAATTAGCTTCCGACTC ATTCATCCAAATACCGAGTCGCCAAAATTC
15	Pri-miR319a_long_Rev	GGAATACAAAAGAGAGAGGGAGCTCCCTTCAGTCCAATCAAAGAG AATCAATGATCCAATTTGTCTACCGCATCATTCATTCAATTAACGAG



		TCTAGTTTGAATTTTGGCGACTCGGTATTTGG
16	Pri-miR394a_Rev	ACTTGTATATAGATCAAGAGGGTTTACGAAG
17	Pri-miR394a_long_Fw	CTGTAATACGACTCACTATAGGTTTGTTCCTTTTCATGAGGGTTTG ACAAAGATTTTCTTACAGTCATCTTTGGCATTCTGTCCACCTCCTTCT ATACATATATGCATGTGTATATATATATGCGTTTCGTGT
19	Pri-miR394a_long_Rev	ACTTGTATATAGATCAAGAGGGTTTACGAAGAAGCCCTAACAGAT CTCTATTGGCAGTATACCCACCTCCTTCTTTCACACGAAACGCATAT ATATATACACATGCATATATGTATAGAAG


#### 2.6.4. Extracción Fenol / Cloroformo / Isoamílico

Llevamos las muestras a purificar a un volumen final de 100µl con agua miliQ, agregamos 100µl de una mezcla Fenol/Cloroformo/Isoamílico 25:24:1 (Roth), homogeneizamos vigorosamente en un agitador vortex durante 1 min y centrifugamos durante 2 min a 13.000 rpm. Tomamos la fase acuosa y le agregamos 100µl de Cloroformo (Roth), homogeneizamos y centrifugamos de la misma manera. Tomamos la fase acuosa y procedimos a la precipitación con etanol de la misma manera que en el apartado 2.6.2.

#### 2.6.5. Transcripción *in vitro* a escala intermedia

Preparamos la mezcla de reacción **conteniendo 35 µg/ml de molde** (oligos hibridados o plásmido linealizado) en una solución amortiguadora de transcripción 1X [Tris 40 mM (pH 8), DTT 5 mM, espermidina 1 mM, Tritón X-100 0,01 % y **PEG 8000 80 mg/ml**], cada rNTP a una concentración de 4 mM (rA, rC, rG y rU), MgCl<sub>2</sub> 20 mM, BSA 40 µg/ml y 1 U de pirofosfatasa. Iniciamos la reacción con el agregado la enzima T7 polimerasa y la misma continuó por 3 horas a 37°C. Luego, agregamos 50 U de ADNasa e incubamos por 30 minutos a 37°C. Diluimos la mezcla de reacción 8 veces en Tris 20 mM (pH 8), EDTA 10 mM y urea 8M y la sembramos en una **columna de intercambio aniónico "Q-Sepharose"** (GE) equilibrada con la misma solución amortiguadora. Eluimos la muestra con un gradiente de 0 a 1 M NaCl en el tampón. Identificamos las fracciones que contenían ARN por A<sub>260</sub>, y las controlamos por electroforesis en gel de poliacrilamida 5 % p/v con Urea 7 M. Dializamos las fracciones que contenían el transcrito deseado contra 200 volúmenes de H<sub>2</sub>O. Liofilizamos distintas fracciones y las almacenamos a -80°C. Al momento de utilizarlas, las resuspendimos en la solución amortiguadora correspondiente, las calentamos a 95°C durante 5 min y las enfriamos inmediatamente en un baño de hielo y agua para asegurar el plegamiento por hibridación intramolecular.

#### 2.6.6. Transcripción *in vitro* a pequeña escala

Como molde empleamos **10µl** de los fragmentos de ADN obtenidos por **PCR**  se describe en el apartado 2.6.3, esta cantidad corresponde a **ca. 2 ug/ml**. Llevamos a cabo la reacción en una solución amortiguadora de transcripción 1X [Tris 40 mM (pH 8), DTT 5 mM, espermidina 1 mM], cada rNTP a una concentración de 4 mM (rA, rC, rG y rU) y MgCl<sub>2</sub> 20 mM. El volumen final fue de 100 ul. La reacción comenzó con el agregado de la enzima T7 polimerasa y continuó por 6 horas a 37°C. Finalizado este tiempo purificamos los productos a través de geles de poliacrilamida desnaturizantes de la misma manera descrita para los oligos de ADN en el apartado 2.6.1.

#### 2.6.7. Determinación de la concentración de muestras de ácidos nucleicos

Determinamos la concentración de muestras de ARN midiendo su  $A_{260}$  en espectrofotómetros UV-Visible en condiciones desnaturalizantes (urea 8M) para evitar la subestimación por efecto hipocrómico de la doble hebra. Obtuvimos el coeficiente de extinción molar para cada muestra a partir de su secuencia empleando el servidor OligoCalc (Kibbe 2007).

## 2.7. Ensayos de mapeo estructural sobre ARN

### 2.7.1. Marcado de muestras de ARN con $^{32}\text{P}$ en extremo 5'

Para realizar los ensayos de mapeo estructural es necesario primero marcar las muestras de ARN en el extremo 5' con fosforo radioactivo. Para esto tomamos 300 pmoles de ARN y realizamos una desfosforilación empleando la enzima Fosfatasa Antartika (NEB) según las instrucciones del fabricante. Luego realizamos una extracción fenol/cloroformo/alcohol isoamílico como se describe en el apartado 2.6.4 y resuspendimos en 15  $\mu\text{l}$  de agua. Tomamos 5  $\mu\text{l}$  de esta muestra y procedimos a realizar la fosforilación del extremo 5' empleando la enzima Polinucleótido Quinasa (Thermo Scientific) de acuerdo a las instrucciones del fabricante y utilizando como sustrato  $^{32}\text{P}$ - $\gamma$ -ATP. Finalmente purificamos estas muestras por gel de poliacrilamida desnaturalizante como se describe en el apartado 2.6.1. Resuspendimos las muestras en agua estimando una concentración final de 1  $\mu\text{M}$  de acuerdo al porcentaje de precursor marcado de largo completo visualizado en el gel.

### 2.7.2. Mapeo por digestión enzimática

Antes de realizar el tratamiento con las enzimas específicas calentamos las muestras de los precursores a 95°C durante 5 min y enfriamos inmediatamente en un baño de hielo/agua para favorecer el plegamiento por hibridación intramolecular. A estas muestras agregamos solución amortiguadora para concentración final 1X conteniendo Tris 20 mM pH 7,5, KCl 25 mM y  $\text{MgCl}_2$  10 mM, ARNt de *E. coli* 0,1 mg/ml (Sigma). En todos los casos el volumen final de la reacción fue de 10  $\mu\text{l}$  y la cantidad de ARN tratado fue de 1,5 pmoles por reacción. Para la digestión con la nucleasa V1 (Ambion) utilizamos 0,01 U por reacción e incubamos durante 15 min a temperatura ambiente. Para la digestión con la nucleasa S1 (Fermentas) empleamos 0,1 U por reacción e incubamos durante el mismo tiempo y temperatura. Como referencia utilizamos una escalera por hidrólisis alcalina, que consiste en tratar 1,5 pmoles de muestra con 3  $\mu\text{l}$  de  $\text{NaHCO}_3$  200 mM pH 10 en un volumen final de 5  $\mu\text{l}$  y calentar durante 3 min a 95°C. Se obtiene la hidrólisis de la cadena en todas las posiciones posibles, por lo tanto se observa en el gel una banda por cada longitud posible, es decir una banda por nucleótido. Otra referencia que empleamos es la digestión con la enzima nucleasa T1, la cual hidroliza el enlace fosfodiéster en posiciones donde encuentra guaninas desapareadas. Para emplear esta enzima como referencia es necesario desnaturalizar la muestra de ARN de forma tal de que todas las guaninas estén expuestas a la enzima. Para lograr esto tomamos 1,5 pmoles de muestra y los llevamos a 10  $\mu\text{l}$  de volumen final en solución reguladora desnaturalizante consistente en citrato 20 mM pH 5, EDTA 1 mM y urea 7 M. Calentamos a 50 °C durante 5 min y luego agregamos 100 U de nucleasa T1 (Fermentas). Incubamos 15 min a 50 °C. En todos los casos al finalizar el tratamiento recuperamos la muestra de ARN tratada por precipitación con etanol como se describió en el apartado 2.6.2 con la diferencia de que en este caso la resuspensión final del pellet la realizamos en 5  $\mu\text{l}$  de solución de siembra desnaturalizante y sembramos las muestras en geles de secuenciación como se describe en el apartado 2.7.4.

### 2.7.3. Mapeo por hidrólisis inducida por Terbio (3+)

De manera similar al tratamiento con enzimas realizamos la hidrólisis inducida por Terbio (3+). Preparamos las muestras de la misma manera y en la misma solución reguladora que en el apartado anterior. La diferencia se encuentra en que, en lugar de las nucleasas, realizamos el tratamiento agregando  $TbCl_3$  a una concentración final de 10 mM. Incubamos 30 min a temperatura ambiente y luego detuvimos el tratamiento agregando EDTA a una concentración final de 100 mM. Inmediatamente después procedimos a realizar la precipitación con etanol como se describe en el apartado 2.6.2, con la diferencia que realizamos el lavado con etanol 70 % por duplicado para eliminar en mayor medida la gran cantidad de iones presentes en la muestra, lo cual afecta la migración en los geles de acrilamida. Resuspendimos las muestras en 5  $\mu$ l de solución de siembra y visualizamos los resultados sembrando las muestras en geles de secuenciación descritos en el siguiente apartado.

#### 2.7.4. Geles de Secuenciación

Preparamos las muestras para sembrar en geles de secuenciación mezclando con solución de siembra desnaturalizante conteniendo formamida 80%v/v, Xilene cyanol 0,025%p/v az  $\mu$ l de bromo fenol 0,025%p/v, solución reguladora TBE (Tris 100 mM, borato de sodio 100 mM, EDTA 2 mM) y EDTA 50 mM pH 8. Calentamos durante 1 min a 95°C, centrifugamos y sembramos 2,5  $\mu$ l inmediatamente en cada pocillo del gel. Preparación previa del gel: utilizamos geles de 8 y 12 % acrilamida conteniendo urea 7 M. Realizamos la polimerización de los geles en vidrios de 40 cm de largo, con separadores de 0,4 mm y peines de 32 calles. Luego de verter la mezcla de polimerización dentro de los vidrios dejamos incubando durante 30 min. Una vez polimerizados, colocamos los geles en las cubas de electroforesis vertical junto con la solución reguladora de corrida (TBE 1X) y precalentamos aplicando un potencial de 3000 V durante 30 min previo a sembrar las muestras. Luego de sembrar las muestras corrimos las electroforesis aplicando el mismo potencial durante 1 h y 40 min para los geles de 12 % de poli(acrilamida) y 2 h para los geles de 8 %. Finalizada la electroforesis transferimos los geles a papel de filtro y cubrimos con film plástico para secar en secador de geles a 80°C durante 40 min. Finalizado el secado colocamos el gel en contacto con placas sensibles a la radioactividad durante toda la noche. Finalmente escaneamos las placas con un escáner óptico. Cuantificamos la intensidad de las bandas obtenidas con el programa SAFA según las instrucciones de los autores (Laederach et al. 2008).

#### 2.8. Titulaciones por anisotropía de la fluorescencia

Para las titulaciones por anisotropía de la fluorescencia, marcamos los fragmentos de ARN en el extremo 5' con fluoresceína usando el sistema 5' EndTag Nucleic Acid End Labeling y la marca reactiva fluoresceína-maleimida de Vector Laboratories. Purificamos los fragmentos marcados con una extracción con fenol, seguida de una precipitación con etanol y resuspendimos en una solución amortiguadora de  $Na_2HPO_4/KH_2PO_4$  10 mM pH7. El ADN marcado con fluoresceína se obtuvo de Sigma. Medimos la anisotropía de la fluorescencia en un espectrofluorómetro Varian Cary Eclipse excitando las muestras a 492 nm y midiendo la emisión de fluorescencia a 520 nm. Los valores de anisotropía se obtuvieron por un promedio de 3 medidas con un tiempo de integración de 20 segundos. La apertura de las rendijas de excitación y emisión fue de 10 nm. Previo a realizar los ensayos calentamos el ARN a 100°C por 5 minutos y enfriando inmediatamente en un baño con hielo y agua para asegurar la hibridación intramolecular. El ADN<sub>dh</sub> lo obtuvimos mezclando las hebras complementarias y calentando a 100°C y dejando enfriar lentamente a temperatura ambiente. Antes de llevar a cabo las titulaciones cambiamos la solución amortiguadora de todas las especies a  $Na_2HPO_4/KH_2PO_4$  10 mM pH 7 utilizando

columnas Bio-Spin 6 (Bio-Rad) de acuerdo a las indicaciones del proveedor. Obtuvimos las curvas de afinidad a temperatura ambiente titulando una solución de ácido nucleico 50 nM marcado, con cantidades crecientes de proteína. Los datos experimentales ( $r$ ) se ajustaron a una función hiperbólica asumiendo un modelo de unión simple con una estequiometría 1:1:

$$r = r_0 + (a * [P]) / (b + [P])$$

Donde  $[P]$  corresponde a la concentración de proteína libre,  $r_0$  es la anisotropía del ARN o ADN libre,  $a$  es la amplitud del cambio de anisotropía tras la unión y  $b$  es la constante de disociación. Normalizamos las curvas de titulación restando el valor de  $r_0$  y dividiendo por la amplitud  $a$  de cada punto.

## 2.9. Dicroísmo circular

Determinamos la elipticidad de las muestras utilizando un espectropolarímetro Jasco 810 calibrado con (+) ácido 10-canforsulfónico. Para la adquisición de espectros en el UV lejano utilizamos una cubeta de paso óptico de 0,1 cm, la solución amortiguadora fue  $\text{Na}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  10 mM (pH7) y la temperatura fue 25 °C. En todos los casos los espectros se adquirieron a una velocidad de 20 nm/min y se promediaron al menos 8 medidas para cada muestra. En los espectros de proteína barrimos el rango entre 250 y 190 nm. En las titulaciones con ácidos nucleicos agregamos alícuotas de soluciones concentradas de proteína a 200  $\mu\text{l}$  de muestras de los distintos ácidos nucleicos, adquiriendo en cada paso un espectro barriendo el rango de 300 a 190 nm. En todos los casos controlamos que el voltaje del fotomultiplicador (HT) no exceda un valor de 600 V. Para analizar las componentes de los puntos de las titulaciones empleamos el análisis por curvas multivariadas (MCR-ALS) (Borges et al. 2005).

## 2.10. Programas y servidores utilizados

MUSCLE (Edgar 2004): método para el alineamiento de múltiples secuencias de proteína.

FastTree (Price et al. 2010): método para construir árboles filogenéticos a partir de secuencias alineadas de proteínas o ácidos nucleicos.

PyMol (Molecular Graphics System, Version 1.3, Schrödinger, LLC): sistema de visualización de moléculas en tres dimensiones. Esta herramienta se utilizó para analizar estructuras y alinear estructuralmente distintas moléculas entre sí.

Resolución por Curvas Multivariadas – Alternando Mínimos Cuadrados (MCR-ALS) (Borges et al. 2005): es una herramienta quimiométrica que permite la resolución de perfiles de concentración y espectros puros de diferentes especies dentro de un sistema con múltiples componentes. Esta herramienta descompone la matriz de los datos experimentales de una titulación en una matriz de evolución de la concentración de los distintos componentes multiplicada por una matriz que corresponde a los espectros de las especies puras más una matriz que contiene los errores experimentales. MCR-ALS es una rutina del programa MATLAB y puede descargarse de <http://www.mcrals.info/>.

TopSpin 2.1 (Bruker): este programa comanda el funcionamiento de los espectrómetros de Bruker utilizados. Se utiliza para diseñar los experimentos, ejecutarlos y visualizar resultados.

NMRPipe (Delaglio et al. 1995): empleamos este programa para procesar los datos crudos (FID) obtenidos de los experimentos de RMN y transformarlos en espectros de frecuencia para poder analizarlos.

CcpNmr Analysis (Vranken et al. 2005): empleamos este programa para el análisis de espectros de RMN, asignación de resonancias, seguimiento de perturbación de señales, y cálculo de desplazamientos químicos secundarios.

SSP (Marsh et al. 2006): Este programa calcula la tendencia a adoptar conformaciones de estructura secundaria tipo hélice alfa, extendidas o lamina beta de proteínas a partir de datos de desplazamientos químicos de los núcleos de carbono alfa, carbono beta y carbonilo del esqueleto peptídico. También proporciona una herramienta para corrección de referenciado de desplazamientos químicos basada en la correlación de los valores de desplazamiento químico de carbono alfa y carbono beta.

Seq2rdc (Huang et al. 2013): este algoritmo calcula el valor de RDC para cada residuo en base a la identidad de los dos vecinos más cercanos, una corrección por ventana de alineamiento local y una corrección de la línea de base que considera la longitud de la cadena.

Flexible-Meccano (Jensen et al. 2008; Bernadó et al. 2005): es un algoritmo diseñado para explorar el espacio conformacional de proteínas desestructuradas de manera eficiente. El muestreo conformacional se basa en la tendencia conformacional de cada tipo de aminoácido y el volumen de las cadenas laterales. A cada residuo se le asigna secuencialmente una combinación de ángulos diedros  $\Phi/\Psi$ . Los valores de los ángulos se seleccionan al azar a partir de una base de datos específica para cada aminoácido. La base de datos fue construida tomando los ángulos diedros de regiones que no forman estructura secundaria regular dentro de estructuras cristalográficas conocidas. El volumen de exclusión de cada tipo de aminoácido es tenido en cuenta para evitar el choque entre los residuos de la cadena. El volumen de cada aminoácido se representa por una esfera (pseudo átomo centroide) a la altura del carbono beta (carbono alfa para las glicinas). En el caso de que se produzca choque estérico el par de ángulos diedros es descartado y se selecciona otro para el mismo residuo.

DSSP 2.0.4 (Kabsch & Sander 1983; Touw et al. 2015): el programa DSSP (por sus siglas en ingles "Define Secondary Structure of Proteins") corresponde al método estándar para la asignación de estructura secundaria para cada aminoácido de una proteína. A partir de las coordenadas de los átomos de una estructura en formato PDB, el programa busca patrones de puentes de Hidrógeno que se correlacionan con los distintos tipos de estructura secundaria. Los tipos de estructura secundaria posibles según DSSP son: hélice alfa (H), hélice  $3_{10}$ (G), hélice  $\pi$  (I), puente beta (B) para regiones de lámina beta extendidas, burbuja beta (E) para interrupciones dentro de láminas beta, giro (T), y curva (S) para sitios en los cuales la dirección de la cadena cambia de forma tal de presentar una curvatura con un ángulo menor a  $70^\circ$ .

SCCOMP (Eyal et al. 2004): este programa modela las cadenas laterales de los residuos sobre una cadena peptídica principal fija. Para el modelado tiene en cuenta superficies de contacto y propiedades químicas de todos los átomos.

SPARTA+ (Shen & Bax 2010): este programa consiste en una red neuronal artificial entrenada para establecer relaciones cuantitativas entre desplazamientos químicos y estructura de proteínas, considerando los efectos de la conformación del esqueleto peptídico y de las cadenas laterales, la presencia de puentes de Hidrogeno y de la corriente de anillos aromáticos. La red neuronal fue entrenada con datos de 580 estructuras cristalográficas de proteínas, para las que se conocen los desplazamientos químicos de los núcleos del esqueleto peptídico y el Carbono  $\beta$ .

Python 2.7 (Python Software Foundation, [www.python.org](http://www.python.org)): es el lenguaje de programación que empleamos para desarrollar las rutinas que nos permitieron construir el ensamble conformacional de DCL1-A.

PyRosetta (Chaudhury et al. 2010): es una interfase para el software de modelado molecular Rosetta basada en el lenguaje de programación Python. PyRosetta permite escribir rutinas personalizadas de predicción de estructuras y diseñar algoritmos usando las principales funciones de muestreo y medición de energía de Rosetta ([www.pyrosetta.org/](http://www.pyrosetta.org/)).

PALES (Zweckstetter & Bax 2000): este programa predice la magnitud y la orientación del alineamiento de un soluto a partir de su forma tridimensional. A partir de la estructura de la molécula el programa simula un tensor de alineamiento asumiendo interacciones puramente estéricas entre la molécula de interés y las partículas del medio de alineamiento. Con el tensor de alineamiento calculado se predicen los valores de RDCs.

CS-Rosetta (Shen et al. 2008): descrito en la sección 2.5.

Mfold (Zuker 2003): servidor que permite la predicción de estructura secundaria de moléculas de ARN a partir de la secuencia (<http://unafold.rna.albany.edu/?q=mfold>).

SAFA v.1.1 (Laederach et al. 2008; Das et al. 2005): del inglés *Semi-Automated Footprinting Analysis*. Este programa consiste en una herramienta desarrollada específicamente para el análisis de los resultados de ensayos de huella sobre moléculas de ácidos nucleicos. Presenta una gran utilidad para la asignación de bandas y cuantificación de intensidades respecto de otros programas con el mismo fin (<https://simtk.org/home/safa>).

### 3. CAPITULO 3: DCL1-A, primera aproximación

DCL1 es una enzima que pertenece a la familia de ARNasas III y junto con las proteínas accesorias HYL1 y SERRATE, lleva a cabo el procesamiento de precursores de miARNs de manera coordinada y precisa en *A. thaliana*. Dentro del complejo de procesamiento, DCL1 es responsable de realizar los cortes en los precursores que dan lugar a los miARN, mientras que HYL1 y SE son proteínas auxiliares cuya función principal es aumentar la precisión de la ARNasa (Dong et al. 2008; Xie et al. 2010).

La secuencia completa de DCL1 se encuentra llamativamente conservada a través de las distintas especies de plantas, en comparación con sus homólogas DCL2, 3 y 4, que presentan una gran variabilidad (Margis et al. 2006). La elevada conservación de la secuencia de esta enzima sugiere que ha alcanzado un grado de refinamiento que probablemente hace posible su rol único y central en la biogénesis de miARNs. En esto se basa nuestro gran interés en el estudio estructural de esta proteína y su mecanismo de funcionamiento.

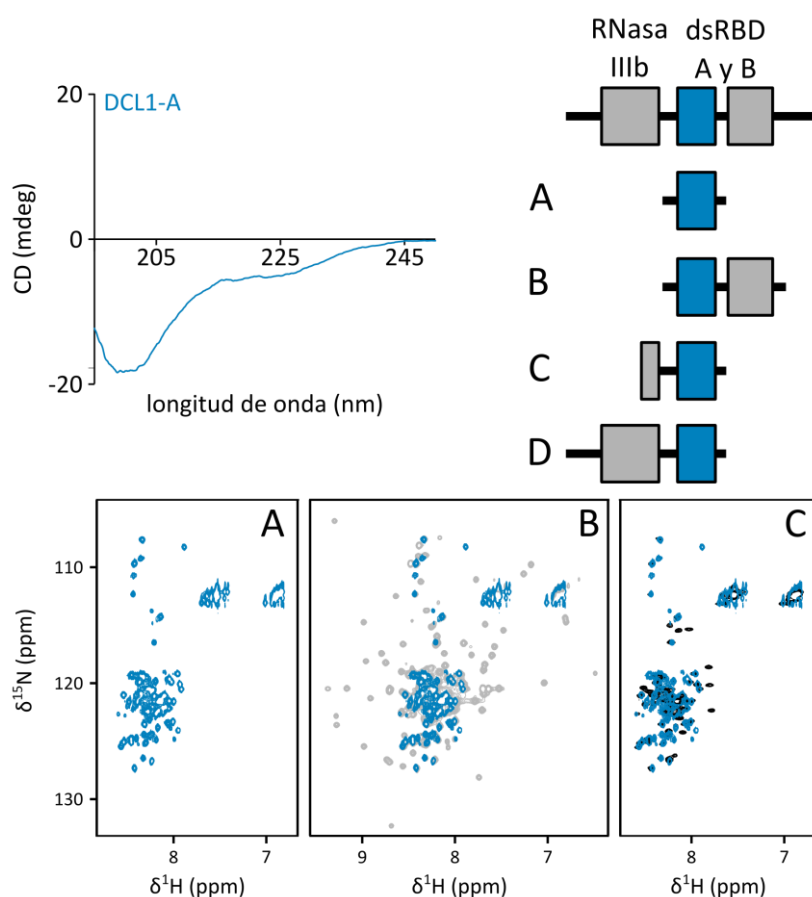
DCL1, posee dos dominios de unión a ARNdH en el extremo C-terminal que en principio podrían conferirle la capacidad de unir los precursores de miARN. La evidencia *in vivo* sugiere que al menos uno de estos dominios es necesario para el funcionamiento de la enzima, ya que una mutante insercional que elimina ambos dominios de la enzima en *A. thaliana* (*dcl1-6*) es letal a nivel del desarrollo del embrión (Schauer et al. 2002). Por otro lado, la mutante *dcl1-9* presenta una inserción que elimina el segundo dsRBD (DCL1-B), dando como resultado el procesamiento impreciso de los precursores de miARN (Schauer et al. 2002). La transformación de las mutantes *dcl1-9* con el fragmento correspondiente a los dos dominios dsRBD en tándem (DCL1-AB) no rescata el fenotipo mutante, indicando que los dsRBDs no pueden cumplir su función en *trans* y son componentes integrales de la enzima (Liu et al. 2013).

Se ha demostrado que DCL1-B es el responsable de la localización de DCL1 dentro del núcleo celular, y más precisamente, de su anclaje a los llamados “cuerpos D” donde se produce la biogénesis de los miARNs (Fang & Spector 2007). En trabajos previos de nuestro grupo se realizó la caracterización biofísica de este dominio, determinando las particularidades estructurales que en principio estarían explicando sus implicancias funcionales (Burdisso et al. 2012).

Acerca del primer dominio dsRBD de DCL1 (DCL1-A), sabemos que es suficiente para el funcionamiento de la enzima, ya que la mutante *dcl1-9* de *A. thaliana* mencionada anteriormente, en la que el único dominio con potencial capacidad de unión al sustrato en el extremo C-terminal es DCL1-A, sigue siendo funcional (Kurihara & Watanabe 2004). En un principio, este indicio nos llevó a tomar la iniciativa de estudiar estructuralmente este dominio y su modo de reconocimiento del sustrato. En un trabajo más reciente, se demostró que cuando este dominio se sobreexpresa aislado no localiza en el núcleo celular ni se han encontrado evidencias para interacciones proteína-proteína con otros componentes del complejo de procesamiento de precursores de miARNs (Liu et al. 2013). En el mismo trabajo, Fang y sus colaboradores demostraron que el doble dominio DCL1-AB es capaz de complementar plantas deficientes en HYL1 (*hyl1-2*). Por paralelismo con las funciones de los dominios de HYL1, este grupo sugiere que la función principal de DCL1-B es la de interacción proteína – proteína, mientras que la de DCL1-A es la de unión al sustrato.

En una primera etapa del trabajo, optimizamos la expresión y purificación del dominio, realizamos una caracterización estructural inicial, comparando su comportamiento con otros dominios homólogos y evaluamos su función como proteína de unión a ARN. DCL1-A es una proteína desestructurada

En el grupo de trabajo disponíamos previamente de una construcción correspondiente al primer dominio dsRBD de DCL1 (DCL1-A), obtenida de bibliotecas de ADN copia de *A. thaliana*. El fragmento corresponde a los residuos M1732-N1811 según la secuencia de la enzima completa (Uniprot ID: Q9SP32). La secuencia se encuentra clonada en un vector pET-TEV para su expresión en *E. coli* fusionada a una cola de Histidina (Kapust et al. 2001). Luego de purificar el dominio aislado en condiciones nativas encontramos que **la proteína soluble expresada presenta espectros de dicroísmo circular y RMN (HN-HSQC) característicos de un polipéptido desplegado**, evidenciado por el mínimo a 200 nm y la baja dispersión de los desplazamientos químicos de los protones amida respectivamente (Figura 3.1).



**Figura 3.1 DCL1-A se encuentra desestructurada aún en presencia de regiones adyacentes de DCL1.**

Arriba izquierda: Espectro de Dicroísmo Circular en el UV lejano de DCL1-A. Arriba derecha: Construcciones ensayadas: DCL1-A, dsRBD-A aislado (A), DCL1-AB, dsRBD-A y B unidos (B), DCL1-AN, incluyendo 25 residuos hacia el dominio RNase IIIb (C), y DCL1-ARb, dsRBD-A unido a RNase IIIb completo (D). Los paneles inferiores corresponden a la superposición del espectro HN-HSQC del dominio aislado (A, azul) con las construcciones B (gris) y C (negro). Se destaca la baja dispersión de los desplazamientos químicos de los protones amida del espectro de DCL1-A, el cual no cambia en el contexto de las otras dos construcciones.



La ausencia de una estructura definida puede deberse a que no estén bien definidos los límites del dominio o que necesite de otras regiones de la proteína para plegarse. Para explorar esta posibilidad clonamos y expresamos distintas construcciones incluyendo regiones hacia los extremos N y C-terminal de DCL1. Los espectros de correlación HN-HSQC de las distintas construcciones muestran que **el dominio permanece desplegado aun en presencia de las regiones vecinas de la proteína** (Figura 3.1). En el panel B se puede ver que el espectro de la construcción DCL1-AB (M1732-S1909, gris) es exactamente la suma de los espectros de los dominios dsRBDs por separado. El espectro negro en el panel C corresponde a una construcción extendida 25 residuos hacia el extremo N-terminal incluyendo parte del dominio ARNasa IIIb (DCL1-AN, G1707-G1810). Se pueden observar algunas señales adicionales a las de DCL1-A aislado que corresponden a los residuos extra incluidos. La construcción D (DCL1-ARb) no se pudo analizar ya que al sobreexpresarla se obtiene en la fracción insoluble y los intentos por replegarla no fueron productivos.

Habiendo determinado que las regiones circundantes no contribuyen a la estabilización de la forma plegada del dominio, decidimos continuar nuestro trabajo utilizando la construcción que corresponde al dominio DCL1-A aislado.

### 3.2. DsRBDs similares a DCL1-A

#### 3.2.1. DsRBDs de estructura conocida

Con la intención de validar nuestros resultados procedimos a buscar una proteína de referencia respecto de la cual pudiéramos definir si el estado desestructurado de DCL1-A es una característica particular de la enzima DCL1 o se repite en dominios homólogos. Para esto realizamos un **análisis de la similitud de secuencia de DCL1-A con los dsRBDs de estructura conocida hasta la fecha**. Empleando el programa MUSCLE (Edgar 2004) realizamos el alineamiento de las secuencias (Figura 3.2, superior). Con este alineamiento construimos un árbol con el programa FastTree (Price et al. 2010), que muestra la cercanía entre los dsRBDs según la similitud entre las secuencias alineadas (Figura 3.2, inferior). De esta forma **logramos identificar a los dsRBDs de Dicer de ratón (ID PDB: 3C4B) y de la levadura *Kluyveromyces polysporus* (ID PDB: 3RV0) como los de mayor similitud de secuencia** (Figura 3.2), siendo esta del 51% y 37% respectivamente.

Ambas estructuras han sido resueltas por cristalografía y se encuentran en el contexto de construcciones de mayor tamaño que incluyen regiones circundantes de las proteínas. En estas construcciones los dsRBDs adquieren un plegamiento correcto. En el caso del dsRBD1 de la enzima Dicer de *K. polysporus* puede inferirse de la estructura cristalográfica que el dominio se pliega como una unidad independiente, ya que no se encuentra en contacto con otras regiones de la proteína y la conexión con el dominio ARNasa III precedente consiste en una región altamente flexible que no presenta densidad electrónica en el cristal (Weinberg et al. 2011).

En cuanto al dominio de la enzima de ratón, la construcción cristalizada corresponde al dominio ARNasa IIIb unido al dsRBD. En la Figura 3.3 puede observarse que ambos dominios se encuentran cerca en la estructura e incluso presentan una superficie de contacto que podría contribuir a la estabilidad del dsRBD (Du et al. 2008). Para definir si esta interacción es esencial para el plegamiento de la proteína y, por ende, podría resultar también importante para la estabilidad de DCL1-A procedimos a expresar, purificar y evaluar la estructura de este dominio aislado en nuestro laboratorio. La secuencia empleada corresponde a la región R1845-Q1910 de

la secuencia anotada de la proteína completa (UniProt ID: Q8R418) y fue seleccionada a partir del análisis de la estructura cristalográfica.

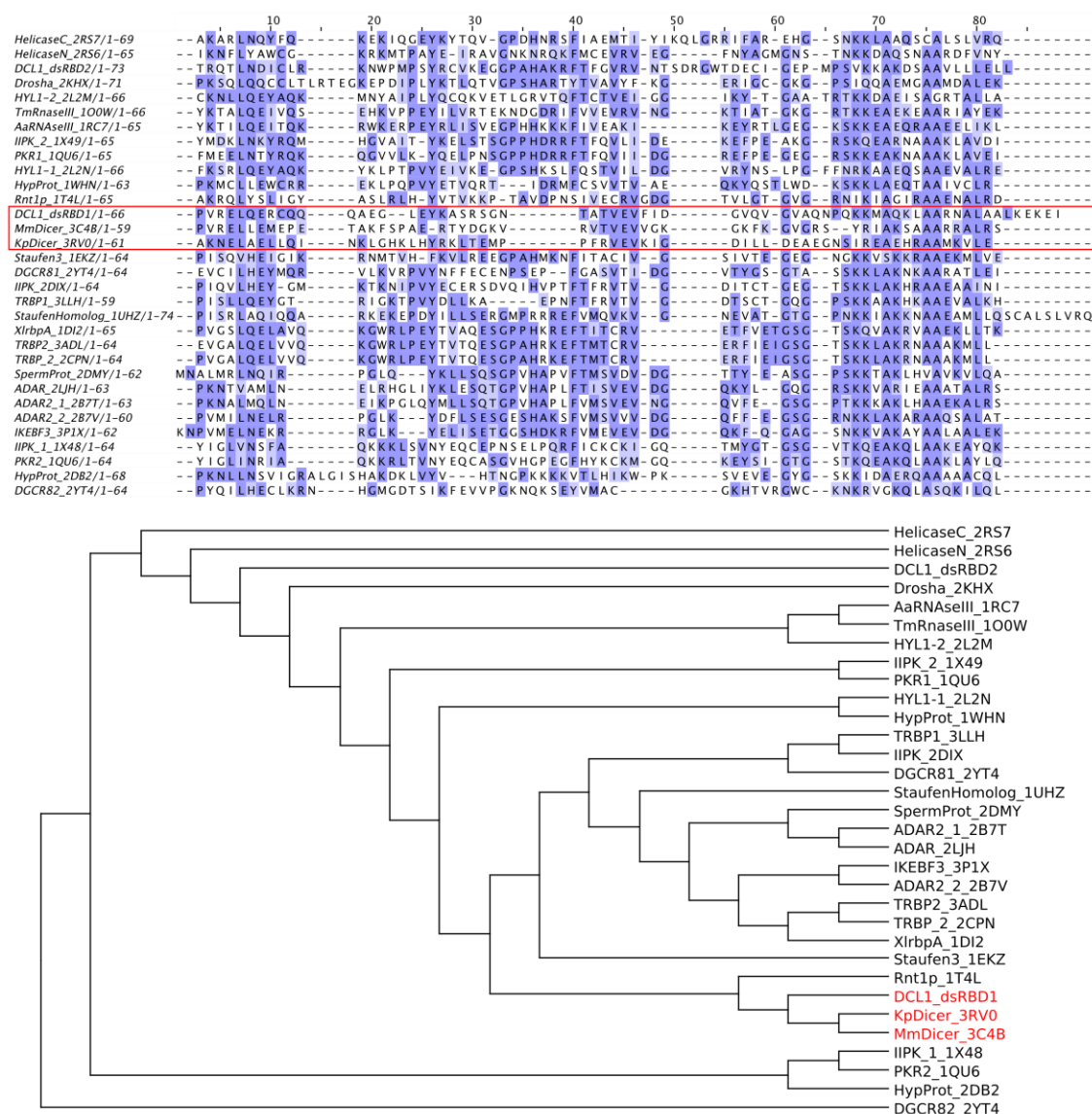
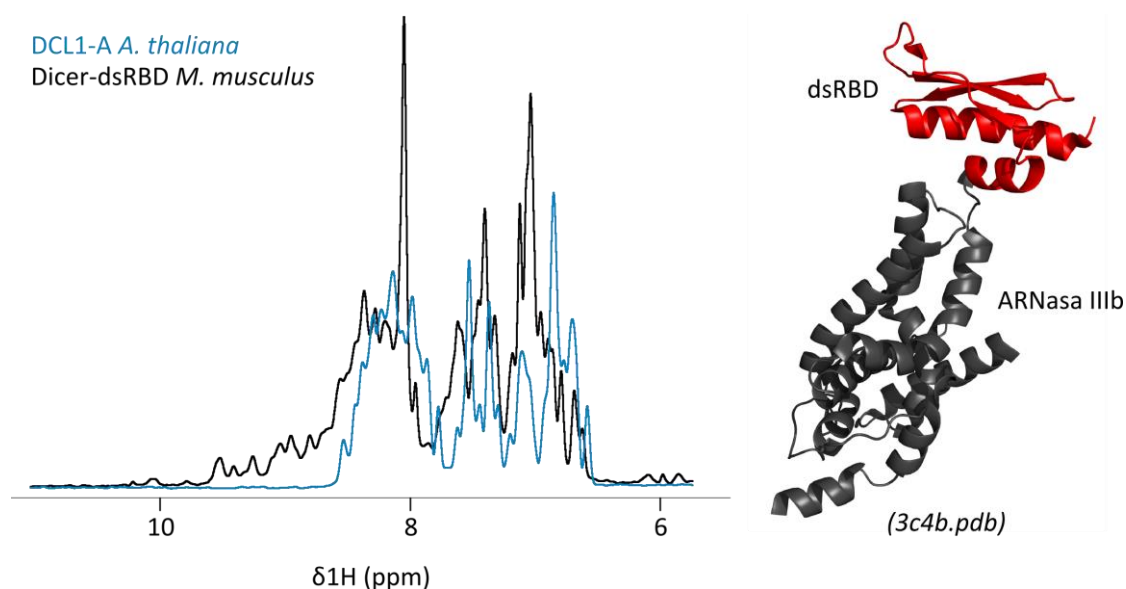


Figura 3.2 Similitud de secuencia de DCL1-A con dsRBDs de estructura conocida.

Superior: secuencias de dsRBDs de estructura conocida alineados con el programa MUSCLE (Edgar 2004). En azul se señala el grado de conservación de los residuos de aminoácido. En rojo se recuadran las secuencias correspondientes a DCL1-A (DCL1\_dsRBD1/1-66), el dsRBD de Dicer de ratón (MmDicer\_3C4B/1-59) y el dsRBD1 de Dicer de *K. polysporus* (KpDicer\_3rv0/1-61). Inferior: árbol de relación entre los dominios construido con FastTree (Price et al. 2010) basado en la similitud de secuencia según los alineamientos de MUSCLE.

A partir de los espectros que se muestran en la Figura 3.3 concluimos que el dsRBD de Dicer de ratón aislado es capaz de adquirir una conformación plegada independientemente del contacto con el dominio ARNasa III. La conformación plegada es la que origina las señales de los protones amida dispersas en el rango entre 8 y 10 ppm (negro). Esta característica de los espectros de polipéptidos estructurados contrasta con el caso de DCL1-A (azul), en el cual estas señales se encuentran concentradas alrededor de 8 ppm.

Habiendo confirmado que el dominio de estructura conocida de mayor similitud de secuencia es capaz de adquirir una conformación plegada independientemente de las regiones circundantes de la proteína, en las mismas condiciones que DCL1-A, podemos concluir que el desorden intrínseco de este último podría ser una característica particular de la enzima DCL1. Teniendo esta hipótesis en mente, decidimos evaluar el grado de plegamiento de los dominios homólogos a DCL1-A de las enzimas tipo Dicer ortólogas a DCL1 en *A. thaliana*, DCL3 y 4.



**Figura 3.3 Dicer-dsRBD de *M. musculus* se pliega como un dominio independiente.**

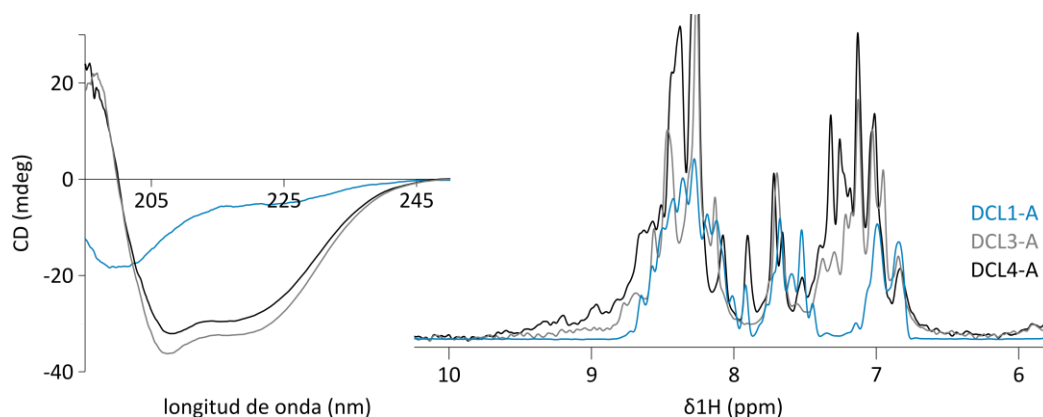
*Izquierda: superposición de los espectros de RMN de protones de los dominios DCL1-A (azul) y Dicer-dsRBD de ratón (negro). Derecha: estructura cristalográfica de la región ARNasa IIIb-dsRBD de Dicer de ratón (PDB ID: 3c4b; (Du et al. 2008)).*

### 3.2.2. DsRBDs de las proteínas tipo Dicer de *A. thaliana*

En todas las especies de plantas multi-celulares encontramos al menos cuatro tipos de enzimas tipo Dicer (Margis et al. 2006), cada uno involucrado en distintas vías de regulación génica a través de ARNs pequeños (Bologna & Voinnet 2014a). Según el árbol filogenético de la evolución de las especies aceptado actualmente, estos cuatro tipos de enzimas se originaron por duplicación génica a partir de dos genes originales, llegando a ser cuatro copias al momento de separarse las plantas superiores de las algas verdes. Estas cuatro copias se conservaron a lo largo de la posterior evolución de las especies de plantas, incluso luego de la separación entre mono- y dicotiledóneas. Particularmente en *A. thaliana* existen cuatro enzimas tipo Dicer, DCL1 a 4. Todas presentan una arquitectura de dominios similar, con la excepción de DCL2, que carece del segundo dsRBD C-terminal. La estructura de los dsRBDs de estas enzimas homólogos a DCL1-A no se conoce. Teniendo en cuenta que el ancestro común entre las DCLs de plantas es más cercano que la distancia entre DCL1 y Dicer de ratón, pensamos que sería interesante determinar si los dsRBDs de las otras DCLs de *Arabidopsis* se pliegan de forma autónoma. De esta forma podríamos confirmar si la falta de estructura de DCL1-A es particular de esta enzima o si se originó más temprano en la evolución de las especies de plantas superiores.

A partir de las secuencias anotadas para DCL3 y DCL4, seleccionamos las regiones correspondientes a los dsRBDs homólogos a DCL1-A por alineamiento de las secuencias. Las proteínas purificadas corresponden a L1365-S1447 para DCL3-A (UniProt ID: Q9LXW7) y I1461-

K1543 para DCL4-A (UniProt ID: P84634). Al realizar la purificación nativa de estas construcciones obtuvimos muestras con una tendencia a precipitar y cuyo espectro de RMN presenta señales anchas, sugiriendo la formación de agregados. Por lo tanto realizamos la purificación de los dominios expresados en condiciones desnaturalizantes y posterior replegado. Sobre las muestras de cada una de estas proteínas adquirimos los espectros de Dicroísmo Circular en el UV lejano y RMN de protones que se muestran en la Figura 3.4.



**Figura 3.4 Estructura de los dsRBD1 de DCL3 y DCL4 de *A. thaliana*.**

*Izquierda: Espectros de Dicroísmo Circular en el UV lejano de DCL1-A (azul), DCL3-A (gris) y DCL4-A (negro). Derecha: región de protones amida de espectros de RMN de las mismas muestras de los espectros de DC.*

Los espectros de Dicroísmo Circular tienen un máximo cercano a 190 nm y dos mínimos, uno alrededor de 210 nm y otro cercano a 220 nm, siendo el mínimo a 210 nm más intenso que el de 220 nm. Estas características corresponden a proteínas con regiones de hélice alfa separadas de láminas beta, lo cual concuerda con lo esperado para un plegamiento tipo dsRBD. Los espectros de dicroísmo circular de DCL3-A y DCL4-A contrastan con el obtenido para DCL1-A, el cual muestra un mínimo pronunciado a 200 nm característico de proteínas desestructuradas. En los espectros de RMN de protones se observan señales hasta alrededor de 9.8 ppm en los casos de DCL3 y 4 indicando que la proteína se encuentra estructurada. Al igual que en el apartado anterior (Apartado 3.2.1, Figura 3.3), vale notar la comparación con el espectro de DCL1-A, en el cual la dispersión de las señales no llega más allá de 8.7 ppm debido a que los protones amida de los enlaces peptídicos se encuentran en entornos muy similares por la conformación mayormente desplegada de la cadena.

Estos resultados estarían indicando que la falta de estructura del dsRBD1 es una particularidad de DCL1 entre las enzimas tipo Dicer de plantas. Vale destacar que la similitud de secuencia entre DCL1-A y sus dominios homólogos en *A. thaliana* no es sustancialmente mayor a la que encontramos con los dominios homólogos de las especies más alejadas analizadas en el apartado anterior (Apartado 3.2.1), siendo 52% con DCL2-A, 44% con DCL3-A y 43% con DCL4-A, por lo tanto no es sorprendente que no compartan las características estructurales.

Al analizar los alineamientos de la secuencia completa de cada tipo de DCL entre distintas especies observamos que la conservación de DCL1 es notoriamente mayor en comparación con las otras tres. En la Figura 3.5 se muestra la región correspondiente al dsRBD1 de cada una de las enzimas. Esto nos lleva a pensar que la falta de estructura de DCL1-A es una característica conservada entre las distintas especies de plantas y por lo tanto es de importancia para la función de la enzima. Nos preguntamos entonces, si podría tener alguna relación con la

diferencia en la función llevada a cabo por las distintas Dicers de plantas. Al respecto, podemos relacionar la diferencia estructural entre DCL1 y las otras tres enzimas con el hecho de que los sustratos de la primera consisten en precursores de miARNs con estructura de hebilla que presentan un gran número de bases desapareadas y burbujas intercaladas en la región doble hebra, mientras que los sustratos de las otras tres enzimas presentan principalmente estructuras de doble hebra perfectas. La flexibilidad del dsRBD1 de DCL1 podría ser una característica que posibilite el reconocimiento de sus múltiples sustratos brindando la capacidad de leer su gran heterogeneidad estructural.

### DCL1-A

gi|15223286|ref|NP\_171612.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALAALEKEETAESEKKE  
gi|225453660|ref|XP\_002268369.1| MHPVREIQ ERCCQQAEGLEYKATRSNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|356504203|ref|XP\_003520888.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALAALEKEETAESEKKE  
gi|357512663|ref|XP\_003626620.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALAALEKEETAESEKKE  
gi|125584735|gb|EAA25399.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|357114316|ref|XP\_003558946.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|255548081|ref|XP\_002515097.1| MHPVREIQ ERCCQQAEGLEYKATRSNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|168010516|ref|XP\_001757950.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|302765743|ref|XP\_002966292.1| MHPVREIQ ERCCQQAEGLEYKASRSGNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|222844405|gb|EEE81952.1| MHPVREIQ ERCCQQAEGLEYKATRSNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE

### DCL2-A

gi|334185051|ref|NP\_001189798.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|357115872|ref|XP\_003559709.1| RDPVREIQ ERCCQQAEGLEYKATRSNLTATVEVFIDGVQVGVAAQNPQKKMAQKLAARNALVLEKEETAESEKKE  
gi|189908862|gb|ACE60552.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|356533095|ref|XP\_00355104.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|357448551|ref|XP\_003594551.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|115453927|ref|NP\_001050564.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|224101241|ref|XP\_002312197.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|255547506|ref|XP\_002514810.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|359476600|ref|XP\_002269915.2| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS  
gi|242033791|ref|XP\_002464290.1| LHPVREITELCQKQFELSKAKD-FDS---FTYEVKAKKMSFAHT-AKASDKKMAKKLAVKEVNLKNS

### DCL3-A

gi|334185717|ref|NP\_001154662.2| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|330370549|gb|AEC12444.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|296081735|emb|CB120740.3| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|224141703|ref|XP\_002324204.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|255550219|ref|XP\_002516160.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|356509849|ref|XP\_003523657.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|357465903|ref|XP\_003603236.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|115441737|ref|NP\_001045148.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK  
gi|242059623|ref|XP\_002458957.1| LPPVREINELCDSLGYPFRVKSNN--DGVKAAQT IQLODDVLLTGDDGSEQTNNLL LGRKAASH LUTQLEK

### DCL4-A

gi|145358264|ref|NP\_197532.3| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|255550579|ref|XP\_002523532.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|359484756|ref|XP\_002264486.2| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|356546104|ref|XP\_003541471.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|224088236|ref|XP\_002308384.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|357167959|ref|XP\_003581414.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|251764670|sp|ATL26.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK  
gi|157285013|gb|ABV31246.1| FIKELINLCQSYKWDREIS--ATKKGDAETWELKVTNKGCLTVSTGRNKRKGTAKKAO LMTNLAHENTITSHPTDVK

Figura 3.5 La secuencia de DCL1 se encuentra más conservada que la de sus parálogos.

En rojo se recuadra la secuencia del dsRBD de la enzima tipo Dicer correspondiente de *A. thaliana*. Se muestran alineamientos con secuencias del mismo tipo de Dicer de otras especies (*Brachypodium distachyon*, *Brassica rapa*, *Glycine max*, *Medicago truncatula*, *Oryza sativa*, *Populus trichocarpa*, *Ricinus communis*, *Vitis vinífera*, *Sorghum bicolor*). En azul se señala el grado de conservación de cada residuo.

Con estos resultados concluimos que el dsRBD1 de DCL1 de *A. thaliana* libre en solución se encuentra en una forma desestructurada. Esta característica es peculiar de este dominio y lo diferencia de otros dsRBDs parálogos. Teniendo esto en consideración continuamos nuestro trabajo con la construcción que consiste sólo en el dominio aislado, ya que no se justifica incluir regiones adyacentes. El siguiente paso fue probar si, a pesar de su falta de estructura, DCL1-A es capaz de unir ARN doble hebra.

### 3.3. Afinidad por el sustrato

Dado que la proteína se encuentra desplegada, nos propusimos determinar si la afinidad por el ARNdH es comparable a la de un dsRBD estructurado.

Para estudiar el equilibrio de unión con el ARNdh realizamos curvas de titulación en las que a una solución de ARN marcado con Fluoresceína en el extremo 5' agregamos cantidades crecientes de proteína. La formación del complejo hace que aumente el tiempo de correlación rotacional de la sonda fluorescente dando como resultado un aumento en la anisotropía de la fluorescencia. El cambio en la anisotropía hace posible monitorear la titulación para finalmente calcular la constante de disociación del complejo.

Realizamos titulaciones de dos fragmentos de ARN diferentes, uno correspondiente a la región del dúplex miARN-miARN\* y otro a la región del tallo inferior (pri-miR172a-ti), ambos del precursor de miR172a de *A. thaliana* (Figura 3.6).

Elegimos trabajar con el fragmento del tallo inferior debido a que su estructura secundaria es de gran importancia para el correcto procesamiento del precursor (Mateos et al. 2010). Este fragmento presenta varias imperfecciones en el apareamiento de bases. Este tipo de imperfecciones está presente en menor proporción en el fragmento que incluye al dúplex miR172-miR172\*, lo cual los hace esencialmente diferentes en cuanto al reconocimiento de su estructura secundaria por parte de un dsRBD.

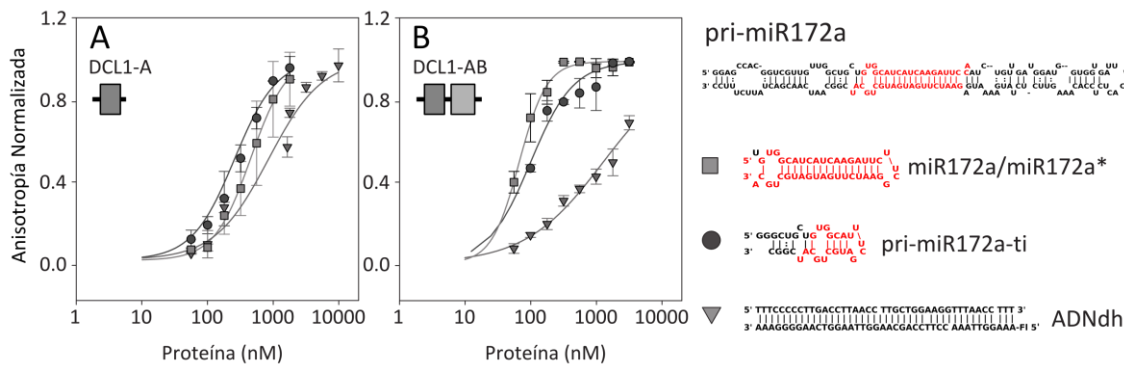
El resultado de estos ensayos se muestra en la Figura 3.6. El dominio **DCL1-A, a pesar de estar desestructurado, une ARNdh con afinidades similares a las presentadas por otros dominios tipo dsRBD**. Las Kd para los fragmentos estudiados fueron de  $723 \pm 250$  nM para el dúplex miR172a/miR172a\* y  $300 \pm 50$  nM para el tallo inferior. La pequeña diferencia de afinidades sugiere que el dominio no estaría prefiriendo un tipo de estructura secundaria por sobre otro. Los valores obtenidos se resumen en la Tabla 3.1.

Como control negativo realizamos una titulación sobre un fragmento de ADNdh que se encontraba disponible en el laboratorio. Este fragmento consiste de dos oligonucleótidos simple hebra hibridados, un oligonucleótido marcado con Fluoresceína en el extremo 5' y un oligonucleótido perfectamente complementario sin marca. Sorprendentemente, encontramos que **la afinidad de DCL1-A por el ADNdh es similar a la que presenta por el ARNdh** (Tabla 3.1), contrariamente a lo esperado para dominios de esta familia que son capaces de diferenciar entre ARNdh, ARNsh y ADN. Obtuvimos un resultado similar con el segundo dominio dsRBD de DCL1 (DCL1-B). En este caso atribuimos la versatilidad de DCL1-B a la alta flexibilidad de la cadena en la región correspondiente al extremo C-terminal de la hélice  $\alpha 1$ , donde se encuentra la región 1 de interacción con el sustrato en los dsRBDs canónicos (Burdisso et al. 2012). Por tratarse de una cadena altamente flexible, podemos aplicar este mismo razonamiento para DCL1-A.

Como control de nuestros ensayos determinamos paralelamente la afinidad del dsRBD1 de la proteína accesoria HYL1 (HYL1-A), un dsRBD canónico (Rasia et al. 2010), por los mismos ácidos nucleicos en las mismas condiciones. La constante de disociación para el complejo con ARN está dentro de los valores esperados, mientras que la del complejo con ADN no pudo ser determinada por ser mayor a  $20 \mu\text{M}$  (Burdisso et al. 2012). Este resultado avala la metodología empleada para determinar las afinidades de los complejos con los dominios de DCL1.

Dado que se ha demostrado que los dsRBDs en tándem generalmente dan lugar a un aumento de la afinidad por el sustrato o especificidad en la unión respecto a los dominios aislados (Acevedo et al. 2015), repetimos estos mismos experimentos con la construcción DCL1-AB, que consiste en los dos dominios dsRBD de DCL1 unidos por el conector silvestre de la enzima. Las constantes de disociación obtenidas fueron de  $67 \pm 9$  nM para el miR172a/miR172a\*,  $100 \pm 11$  nM para el pri-miR172a-ti y  $> 1 \mu\text{M}$  para el ADNdh.

Comparando los valores resumidos en la Tabla 3.1, concluimos que la afinidad por el ARNdh por parte del doble dominio es algo mayor que la de los dominios aislados, siendo esta diferencia más marcada para el dúplex miR172a/miR172a\* que para el fragmento pri-miR172a-ti. Este aumento es esperable por ser que la construcción DCL1-AB presenta dos sitios de unión al sustrato que contribuyen a aumentar la afinidad. La diferencia en el aumento relativo de la afinidad por los diferentes sustratos (10 veces para el miR172a/miR172a\* contra 3 veces para el tallo inferior) podría atribuirse a que el fragmento miR172a/miR172a\* es tres pares de bases más largo que el pri-miR172a-ti, lo cual puede ser de menor importancia para la unión de los dominios aislados, pero puede influir en la unión del doble dominio por aumentar los impedimentos estéricos entre los dominios, resultando en un menor aumento en la afinidad relativa. Por otro lado es notoria la gran diferencia entre la afinidad por el ARNdh y el ADNdh por parte del doble dominio en comparación con los dominios aislados. Este resultado resulta interesante por ser el ARN, y no el ADN, el sustrato esperado para esta región de la enzima.



**Figura 3.6** Determinación de la afinidad por el sustrato por anisotropía de la fluorescencia.

A. titulación con DCL1-A sobre miR172a/miR172a\* (cuadrados), pri-miR172a-ti (círculos) y ADNdh (triángulos). B. titulación de DCL1-AB sobre los mismos sustratos.

**Tabla 3.1** Constantes de disociación determinadas por anisotropía de la fluorescencia

	miR172a/miR172a*	Pri-miR172a-ti	ADNdh
DCL1-A	723 ± 250 nM	300 ± 50 nM	860 ± 270 nM
DCL1-B <sup>†</sup>	810 ± 180 nM	350 ± 20 nM	600 ± 50 nM
DCL1-AB	67 ± 9 nM	100 ± 11 nM	> 1 μM
HYL1-A <sup>†</sup>	ND	Ca. 310 nM	> 20 μM

<sup>†</sup>(Burdisso, P., Suarez, I.P., Bologna, N.G., Palatnik, J.F., Bersch, B. & Rasia 2012)

Habiendo establecido que DCL1-A es capaz de formar un complejo con el ARNdh y que participa junto a DCL1-B en la selectividad por el sustrato, a pesar de encontrarse desplegada, continuamos nuestro estudio determinando la conformación adquirida por esta proteína en presencia del ligando.

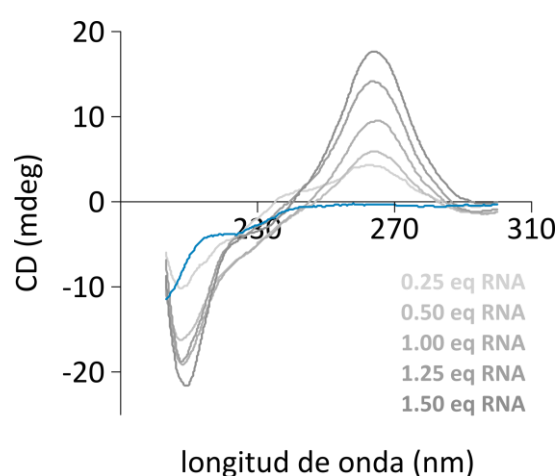


### 3.4. DCL1-A se pliega al interaccionar con el sustrato

Considerando que la unión de proteínas desestructuradas a ácidos nucleicos puede producirse a través de interacciones electrostáticas inespecíficas o por adquisición de una conformación plegada, decidimos evaluar si DCL1-A se pliega en presencia de su sustrato.

#### 3.4.1. Plegamiento por unión a ARNd<sub>h</sub>

Para estudiar la adquisición de estructura secundaria en presencia de ARN realizamos agregados secuenciales de una solución concentrada del fragmento pri-miR172a-ti a una dilución de DCL1-A. Para cada punto registramos un espectro de dicroísmo circular en la región del UV lejano. El resultado se muestra en la Figura 3.7. El espectro azul es típico de un polipéptido desplegado y corresponde a la proteína libre. Los espectros en escala de grises corresponden a los sucesivos agregados de pri-miR172a-ti.



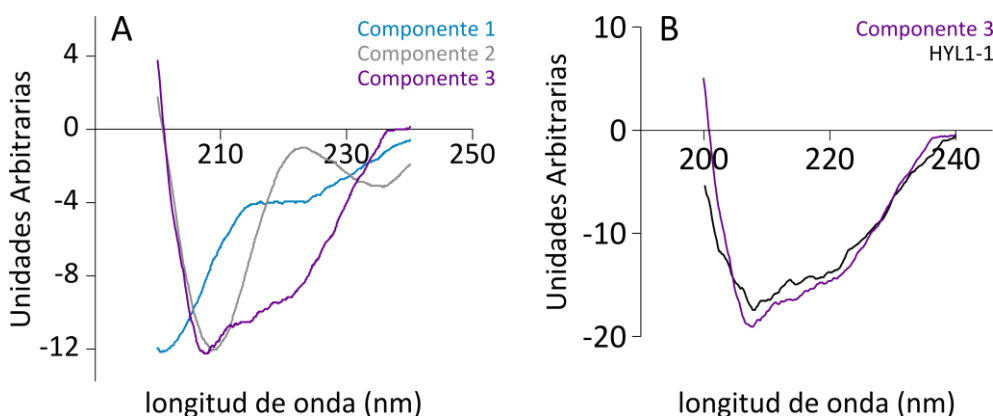
**Figura 3.7** Titulación de DCL1-A con pri-miR172a-ti monitoreada por dicroísmo circular.

*Titulación de DCL1-A con cantidades crecientes de pri-miR172a-ti en escala de grises. El espectro de la proteína libre se muestra en azul.*

El espectro de CD del ARN presenta un mínimo muy intenso alrededor de 210 nm que coincide con una de las bandas del grupo amida del esqueleto de las proteínas. Por esta razón no es posible analizar los espectros de la Figura 3.7 de manera directa, ya que pequeños agregados de pri-miR172a-ti enmascaran por completo el espectro de la proteína y, en caso de haber interacción, el análisis de los espectros diferenciales de las especies aisladas introduciría errores. Por esto es que para estudiar las especies presentes en la mezcla durante la titulación recurrimos a una descomposición por resolución de curvas multivariadas (Borges et al. 2005). Este análisis arrojó como resultado tres componentes, una correspondiente al espectro de la proteína desplegada (azul), otra correspondiente al del ARN (gris) y una tercera que se corresponde con el de una proteína con estructura secundaria tipo alfa + beta (púrpura) (Figura 3.8, A). El espectro de la tercera componente es muy similar al del dsRBD1 de HYL1 (HYL1-A), un dsRBD canónico (Figura 3.8 B, (Rasia et al. 2010)). Por lo tanto podemos inferir que muy probablemente DCL1-A está adquiriendo un plegamiento tipo dsRBD cuando entra en contacto con el sustrato. Podemos llevar esta observación a términos semicuantitativos calculando la relación de la intensidad del mínimo cercano a 210 nm respecto del valor a 220 nm, donde se encuentran las bandas de absorción del enlace peptídico que reportan estructura secundaria. El



valor de esta relación para DCL1-A en presencia del sustrato (0,76) es muy similar al de HYL1-A (0,80), de acuerdo con lo esperado (Tabla 3.2).



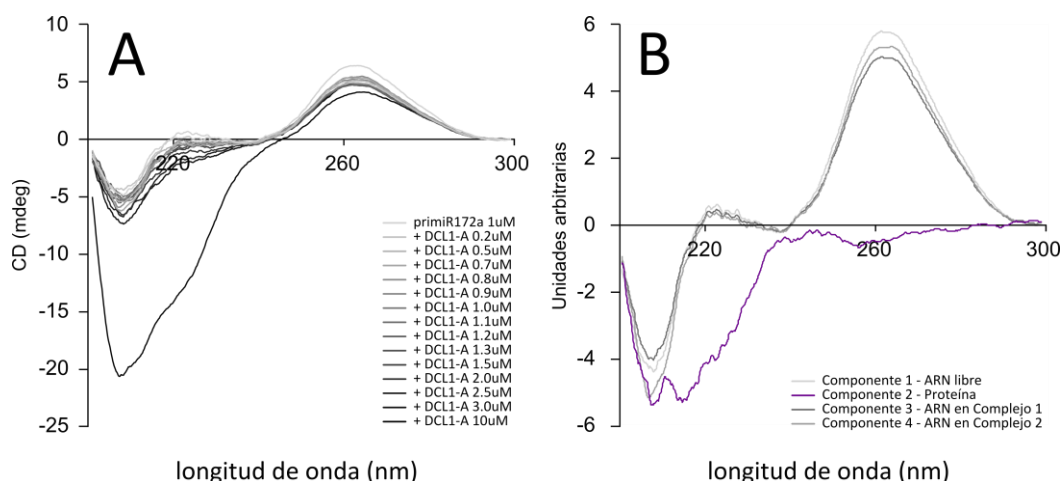
**Figura 3.8 DCL1-A adquiere un plegamiento tipo dsRBD en presencia del sustrato.**

A. Espectros base componentes de la serie de datos de la titulación de DCL1-A con pri-miR172a-ti, obtenidos por análisis de curvas multivariadas (Borges et al. 2005). Azul proteína libre, gris ARN y purpura proteína plegada. B. Comparación de la tercera componente con el espectro obtenido experimentalmente del dominio dsRBD canónico HYL1-A. El eje “y” se muestra en unidades arbitrarias por tratarse de los componentes base de los espectros de la titulación.

Estos resultados estarían sugiriendo que DCL1-A adquiere estructura secundaria en presencia de su sustrato. En el capítulo 5 procederemos a un análisis a mayor resolución de esta estructura empleando resonancia magnética nuclear.

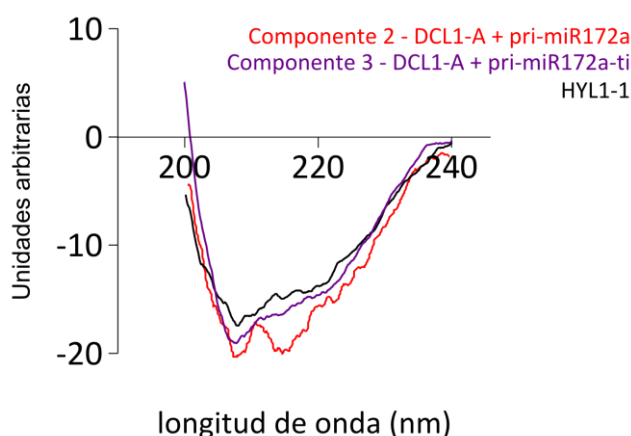
El fragmento del tallo inferior del precursor del miR172a tiene 13 pares de bases de longitud, el tamaño mínimo de ARN doble hebra necesario para la unión de los dsRBDs en general (Ucci et al. 2007). Decidimos corroborar que la estructura adquirida por DCL1-A unida a este sustrato corto es equivalente a la que presenta en presencia de un sustrato de mayor longitud, empleando una metodología similar a la anterior. Para esto realizamos una titulación agregando cantidades crecientes de DCL1-A a una muestra 1  $\mu$ M del precursor completo del miR172a, pri-miR172a, que posee un tallo doble hebra de 67 pares de bases interrumpido por algunas imperfecciones. Los espectros obtenidos se muestran en la Figura 3.9. Inesperadamente, agregando hasta 10 equivalentes de proteína por ARN no logramos ver el espectro de la forma libre desplegada, lo cual nos indica que la estequiometría de unión de DCL1-A al precursor completo es mayor a 10 moléculas de proteína por cada molécula de ARN. Esto nos estaría sugiriendo que el dominio DCL1-A se une por igual a lo largo de todo el precursor, sin seleccionar ninguna región en particular. Al realizar la descomposición de los datos por análisis de curvas multivariadas encontramos cuatro componentes, tres de las cuales coinciden con espectros de ARN y la cuarta corresponde al de una proteína plegada. Respecto a las tres componentes de ARN podemos pensar que se trata de una componente de ARN libre, y otras dos correspondientes a dos formas de estructura levemente diferentes en complejo con la proteína.

La componente correspondiente a la proteína plegada es muy similar a la obtenida del análisis de la titulación con el fragmento más corto de ARNdh (Figura 3.10), por lo tanto concluimos que **la estructura adquirida por DCL1-A es muy probablemente la misma, independientemente de la longitud del sustrato**. Esta observación la corroboramos nuevamente calculando la relación entre las intensidades a 210 y 220 nm (Tabla 3.2).



**Figura 3.9** Titulación de DCL1-A sobre el precursor completo del miR172a monitoreada por CD.

A. Espectros de los distintos puntos de la titulación desde cero a diez equivalentes de proteína agregada sobre 1  $\mu$ M ARN, en escala desde gris hasta negro. B. Componentes presentes en los espectros de la titulación obtenidas por descomposición de curvas multivariadas (Borges et al. 2005). Tres componentes coinciden con espectros de ARN (grises) y la cuarta corresponde al de una proteína plegada (purpura).



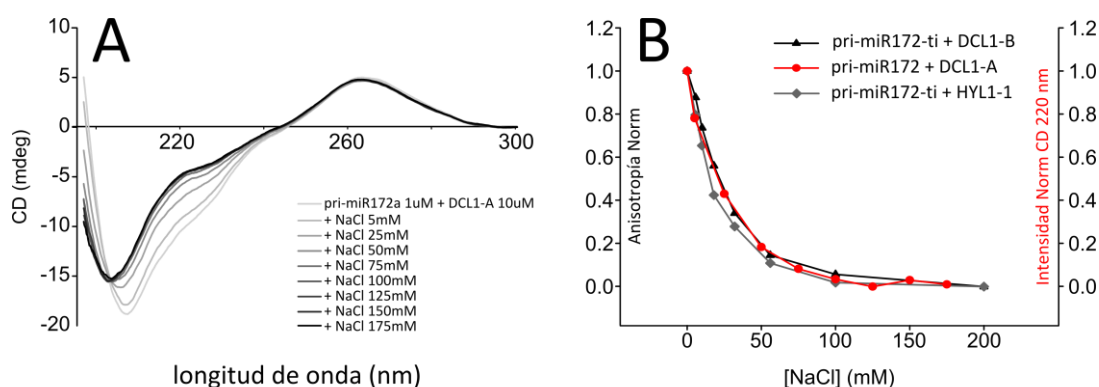
**Figura 3.10** Comparación de las formas plegadas de DCL1-A unida a sustratos de distinta longitud.

Espectro resultante del análisis por descomposición de curvas multivariada de la titulación de pri-miR172a completo (67 pb) (rojo), y la titulación con pri-miR172a-ti (13 pb) (purpura). Como referencia se incluye el espectro experimental de HYL1-A (negro).

**Tabla 3.2** Relación entre la intensidad de las bandas de CD a 220 nm y 210 nm

	CD 220 nm / CD 210 nm
Componente 3, DCL1-A + pri-miR172a-ti	0,76
Componente 2, DCL1-A + pri-miR172a	0,76
HYL1-A	0,80

Por otra parte nos propusimos determinar si el plegamiento de DCL1-A en presencia de ARN es un fenómeno reversible. Para esto decidimos titular con Cloruro de Sodio una muestra de proteína en complejo con ARN, con la hipótesis de que el incremento de la fuerza iónica provocaría la disociación del complejo, liberando la proteína y por ende haciendo visible el espectro de la forma desplegada. Procedimos agregando concentraciones crecientes de Cloruro de Sodio a una muestra de DCL1-A 10  $\mu$ M + pri-miR172a 1  $\mu$ M, registrando un espectro de CD para cada punto. Los espectros se muestran en la Figura 3.11, A. Efectivamente, con el aumento de la fuerza iónica observamos una disminución en la intensidad a 220 nm y un corrimiento del mínimo desde 210 nm hacia longitudes de onda más bajas, claros indicativos de que la proteína está perdiendo su estructura, muy probablemente por la disociación del ARN. Al graficar la evolución de la intensidad de la señal a 220 nm con el agregado de Cloruro de Sodio (Figura 3.11, B, rojo) notamos que la disociación se produce a fuerza iónica relativamente baja, lo cual nos indica que **la interacción depende en gran medida de componentes electrostáticos**.



**Figura 3.11 Reversibilidad del plegamiento de DCL1-A.**

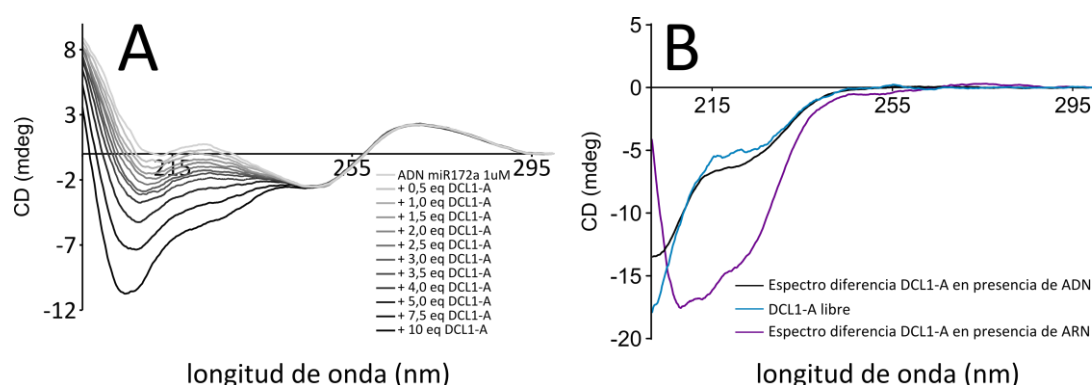
A. Desde gris a negro: agregados secuenciales de Cloruro de Sodio a una muestra con 10  $\mu$ M DCL1-A + 1  $\mu$ M pri-miR172a. B. Evolución de la intensidad normalizada de la señal a 220 nm del complejo DCL1-A:pri-miR172a con el agregado de Cloruro de Sodio (rojo) comparada con la evolución de la anisotropía de la fluorescencia normalizada de una muestra de DCL1-B:pri-miR172a-ti (negro) y de una muestra de HYL1-A:pri-miR172a-ti (gris).

Lo mismo observamos para el segundo dsRBD de DCL1 cuando titulamos con NaCl una muestra de pri-miR172a-ti saturado con DCL1-B, monitoreado por anisotropía de la fluorescencia (Figura 3.11, B, negro). Empleando la misma metodología estudiamos la dependencia de la fuerza iónica de la disociación de HYL1-A del ARNdh (Figura 3.11, B, gris). La disociación de ambos complejos evoluciona de la misma manera con el aumento de la fuerza iónica en comparación con la disociación del complejo de DCL1-A. Esto podría estar sugiriendo que el componente electrostático de la interacción con el sustrato de la especie plegada de DCL1-A es similar al de los dsRBDs canónicos.

#### 3.4.2. La unión de DCL1-A a ADN no induce su plegamiento

Durante nuestros ensayos de afinidad por el sustrato encontramos, inesperadamente, que DCL1-A tiene una afinidad similar por el ADNdh y por el ARNdh. Esto nos sorprendió dado que los dsRBDs han sido caracterizados por su capacidad de distinguir entre ARNdh, ARNsh, y ADN, en base a determinantes estructurales de los distintos tipos de hebras. Estimulados por el hecho de que es posible visualizar el plegamiento de DCL1-A al unirse a ARNdh por Dicroísmo

Circular, procedimos a probar si el mismo fenómeno se da con la unión a ADN. Realizamos una titulación seguida por CD de un fragmento de ADN doble hebra de secuencia correspondiente al miR172a (20 pb) a una concentración de 1  $\mu\text{M}$  con 0.5 a 10 equivalentes de DCL1-A (Figura 3.12, A). De la poca variabilidad de la banda a 270 nm podemos concluir que la estructura del ADN no cambia con el agregado de la proteína. Al sustraer el espectro del ADN libre del espectro del último punto de la titulación, donde DCL1-A se encuentra a una concentración 10  $\mu\text{M}$ , observamos que el espectro diferencia obtenido no coincide exactamente con el de la proteína desplegada (Figura 3.12 B, espectro negro respecto al azul), pero presenta una gran similitud. El análisis por curvas multivariadas de esta titulación no arroja como resultado la presencia de una tercera especie. Concluimos entonces que **la unión de DCL1-A al ADN no resulta en el plegamiento del dominio**, como ocurre con la unión al ARNdh (Figura 3.12, B, púrpura).

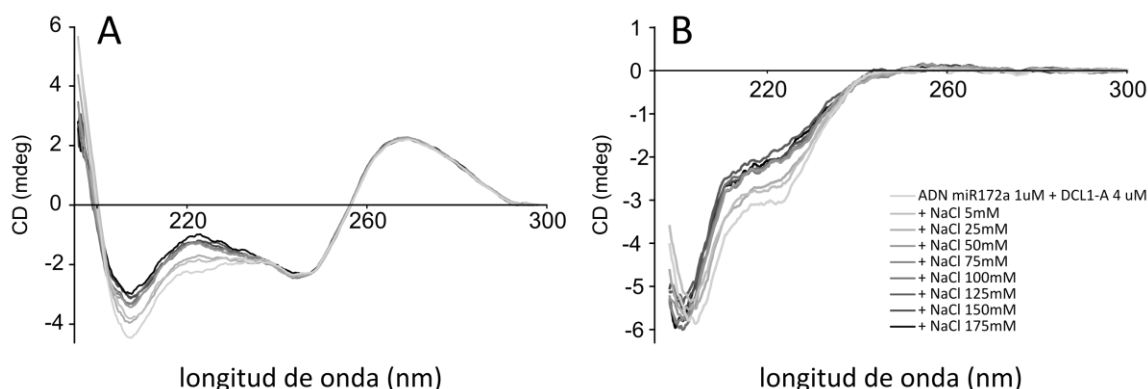


**Figura 3.12 Dicroísmo Circular de DCL1-A en presencia de ADN.**

A. Titulación de concentraciones crecientes de DCL1-A sobre una muestra 1  $\mu\text{M}$  de ADN-miR172a (desde gris claro a negro). B. Comparación de la forma libre de DCL1-A (azul) con la forma en presencia de ADN (negro) (espectro diferencia entre el último punto de la titulación en A, negro, con el espectro del ADN libre, gris claro en A) y la forma en presencia de ARNdh (púrpura). Se destaca la pequeña diferencia de la forma en presencia de ADN (negro) con la forma libre desplegada de DCL1-A (azul), y la gran diferencia de ambas especies con la forma de DCL1-A en presencia de ARNdh (púrpura).

A pesar de no adquirir una estructura tipo dsRBD al unirse al ADN, observamos un pequeño cambio en el espectro de DCL1-A en la titulación anterior. Decidimos confirmar que efectivamente la proteína cambia de conformación en presencia de ADN mediante el estudio de la disociación por aumento de la fuerza iónica de la misma manera que lo hicimos para el complejo con ARNdh en el apartado 3.4.1. Para esto agregamos concentraciones crecientes de NaCl a una muestra de DCL1-A 4  $\mu\text{M}$  + ADN-miR172a 1  $\mu\text{M}$ . Los espectros obtenidos se muestran en la Figura 3.13, A.

Nuevamente en esta titulación observamos que la banda a 270 nm permanece constante, reportando la estabilidad de la estructura de la doble hélice de ADN con el agregado de NaCl. Esto nos permite inferir que los cambios que se observan en los espectros en la región entre 200 y 220 nm se deben a que la forma de DCL1-A unida al ADN presenta una conformación diferente a la forma libre de la proteína. Es posible visualizar este cambio en los espectros diferencia entre cada punto de la titulación y el espectro del ADN libre en la Figura 3.13, B. Como ya observamos previamente esta forma de la proteína unida al ADN es muy diferente a la unida al ARN, y probablemente no presenta un plegamiento definido.



**Figura 3.13** Disociación del complejo de DCL1-A – ADN con el aumento de la fuerza iónica.

*Desde gris a negro: agregados secuenciales de NaCl sobre una muestra de DCL1-A 4  $\mu$ M + ADN-miR172a 1  $\mu$ M. A. Espectros de CD de cada punto de la titulación. B. Espectros diferencia entre cada punto y el espectro del ADN libre. Notar el cambio en los mínimos a 200 y 220 nm por el cambio de conformación en DCL1-A a medida que una mayor proporción de la proteína se disocia del complejo con el ADN por el aumento en la fuerza iónica.*

Dada la elevada flexibilidad de la cadena de DCL1-A no resulta difícil pensar en que los sitios a través de los cuales la proteína interacciona con los ácidos nucleicos se encuentran disponibles para unirse a moléculas blanco muy diferentes, como son en este caso la forma B de la doble hélice del ADN y la forma A de la del ARNdh. Finalmente evidenciamos que la unión a ácidos nucleicos no es condición suficiente para el plegamiento de la topología compleja tipo dsRBD, sino que es necesario que la molécula unida presente sitios de reconocimiento en una geometría específica (hélice A del ARNdh) que permita el reacomodamiento de la cadena de DCL1-A de forma de adquirir la estructura final.

### 3.5. Discusión

En este capítulo definimos el sistema que seguiremos describiendo en los próximos dos capítulos. Determinamos que DCL1-A se encuentra desplegada en solución, independientemente de las regiones de la proteína completa que la rodeen. Por comparación con los dominios homólogos de otras enzimas Dicer, concluimos que la elevada flexibilidad de la cadena parecería ser una particularidad de la enzima DCL1. Dado que este dominio es actualmente el principal candidato a ser el responsable de la unión al sustrato, pensamos en la diferencia entre los sustratos de las distintas Dicer. DCL1 procesa precursores con estructura de hebilla, producto de la ARN polimerasa II. La estructura del tallo de los precursores es altamente heterogénea, presentando una cantidad variable de bases desapareadas y burbujas (Bologna et al. 2009; Xie et al. 2010). Las DCL2, 3 y 4 generan ARNs pequeños de interferencia a partir de ARNs más largos de distintos orígenes y con funciones diversas, pero los precursores son similares entre sí (Bologna & Voinnet 2014b). En general estos precursores son generados por ARN polimerasas dependientes de ARN, que dan como producto ARN doble hebra perfectamente complementario. En conclusión, los precursores procesados por DCL1 son estructuralmente más heterogéneos que los procesados por las otras Dicers, por ende pensamos que la alta flexibilidad de DCL1-A estaría confiriendo versatilidad a la enzima en la capacidad de unión a sustratos muy diferentes entre sí.

A pesar de estar desestructurada DCL1-A es capaz de efectuar su función esperada, ya que une los sustratos con afinidades similares a las de dsRBDs canónicos (Rasia et al. 2010; Masliah et al.

2013). Adicionalmente, encontramos que el dominio presenta afinidad por el ADN, fenómeno que ya habíamos descripto para DCL1-B (Burdisso et al. 2012). En principio se espera que los dsRBDs presenten afinidades mayores por el ARNdh que por otros tipos de ácidos nucleicos, sin embargo se han descripto previamente algunos dsRBDs capaces de unir ADN (Barraud et al. 2011; Bass et al. 1994). La posibilidad de discriminar entre ARNdh y ADN se origina en dos diferencias principales entre estas moléculas, una estructural y otra química. La estructural consiste en el ancho del surco mayor, de 10 Å en la forma A de la doble hélice del ARNdh y de 17 Å en la forma B del ADN. El ancho del surco mayor es detectado por el motivo KKxAK en el extremo N-terminal de la hélice  $\alpha 2$ . Para que esto ocurra, este motivo debe estar en una conformación particular que permita a las lisinas interaccionar con los grupos fosfato en ambos bordes del surco mayor (Vukovic et al. 2014; Masliah et al. 2013; Stefl et al. 2010). Dada la naturaleza flexible de la cadena de DCL1-A, no esperamos que este determinante permita discriminar entre los distintos ácidos nucleicos. Por otro lado la diferencia química entre el ARNdh y el ADN se encuentra en la presencia de los grupos 2'OH, expuestos en el surco menor de la forma A de la doble hélice del ARNdh. Estos grupos forman puentes de hidrógeno con la región C-terminal de la hélice  $\alpha 1$  y el bucle  $\beta 1$ - $\beta 2$ . Por otro lado hay otros tipos de interacciones que los dsRBDs pueden establecer con su molécula blanco que no son específicas al ARN, principalmente del tipo electrostáticas y de puente de hidrógeno con los átomos de oxígeno de los grupos fosfato que constituyen el esqueleto de los ácidos nucleicos, así como también tipo van der Waals con las bases nitrogenadas. Como demostramos en este capítulo, el componente electrostático implicado en estas interacciones es importante. La conclusión es que la principal explicación para la versatilidad en la unión de DCL1-A a distintos blancos radica en la falta de una estructura rígida. La ausencia de rigidez podría estar dando disponibilidad a numerosos grupos con potencial de interacción con el ligando sin restricciones de distribución en el espacio. Esto es lo que podría resultar en la formación del complejo con el ADN sin inducción del plegamiento.

Estos experimentos sentaron la base para proseguir con la determinación de la estructura adquirida por la proteína al interaccionar con el sustrato. Encontramos de gran interés el hecho de que DCL1-A adquiriera una conformación estructurada en presencia de ARNdh, mientras que se mantenga principalmente desordenada al interaccionar con ADN. El hecho de que el plegamiento del dominio dependa del tipo de molécula con la que interacciona nos lleva a pensar que la clave para el plegamiento se encuentra en el arreglo estructural de los puntos de contacto que la cadena desestructurada encuentra sobre la superficie de la doble hélice del ARNdh. Es decir, DCL1-A lee la información en este arreglo estructural para reordenar su cadena de forma tal que se produzcan las interacciones necesarias para adquirir la topología compleja correspondiente al plegamiento tipo dsRBD. Se han descripto varios casos en los que una proteína desestructurada adquiere distintas conformaciones según el ligando al cual se une. Algunos ejemplos son la región regulatoria C-terminal de p53, que puede plegarse en hélice, lámina  $\beta$  o en formas irregulares extendidas dependiendo del complejo en el que se encuentre, así como también el dominio C-terminal activador del factor inducible por hipoxia HIF-1 $\alpha$ , que adopta conformaciones de hélice o extendida dependiendo de la proteína con la que interactúa (Wright & Dyson 2009; Oldfield et al. 2008). Las diferentes estructuras secundarias en estas proteínas son inducidas por la interacción con el ligando a través de distintos grupos de residuos en la misma región.

En el Capítulo 5 profundizaremos la descripción del plegamiento adquirido por DCL1-A y el mecanismo de formación del complejo, empleando resonancia magnética nuclear, técnica que

nos permitirá una aproximación de cuestiones estructurales con resolución a nivel de cada residuo de aminoácido.

Antes de entrar en la caracterización del complejo de DCL1-A con el sustrato nos enfocamos en el estado previo, la descripción del dominio libre en solución. La motivación para estudiar esta especie surge de que uno de los grandes interrogantes acerca de la unión al sustrato y plegamiento de las proteínas desestructuradas se refiere a si es necesaria la existencia de una población de conformaciones parcialmente plegadas previo al contacto con la molécula blanco (Wright & Dyson 2009). Por este motivo durante el siguiente capítulo nos adentramos en la descripción biofísica de la forma libre desplegada ya que consideramos que posiblemente su naturaleza sea de gran importancia para la eficiencia del reconocimiento del sustrato y plegamiento de la proteína.


## 4. CAPITULO 4: ¿Cuán desplegada se encuentra DCL1-A libre?

Como describimos en el capítulo anterior, en su estado nativo, DCL1-A libre se encuentra mayormente desestructurada. Esta característica nos lleva a revisar algunas cuestiones inherentes a la vasta diversidad en grados de plegamiento de las proteínas en general.

El grado de plegamiento de las proteínas consiste en un espectro continuo que varía desde aquellas que presentan estructuras estables, pasando por los glóbulos fundidos, proteínas estructuradas localmente y finalmente las proteínas intrínsecamente desordenadas (IDPs) (Uversky 2013). Con el descubrimiento de que una fracción inesperadamente alta del genoma humano codifica para proteínas funcionales, o dominios de proteínas, que carecen de una estructura terciaria estable comenzó a cuestionarse el concepto de que una única estructura tridimensional es determinante para la función de las proteínas. Esto originó un cambio en el paradigma clásico de la relación estructura – función de las proteínas (Jensen et al. 2014).

Análisis bioinformáticos de diversos proteomas predicen que las IDPs son prevalentes en todos los organismos vivos, con una abundancia particular en proteomas eucariotas. Las IDPs juegan roles cruciales en numerosos procesos bioquímicos, incluyendo transducción de señales, regulación de la función de sus moléculas blanco, y reconocimiento molecular en el ensamble de complejos supra-moleculares (Uversky 2013). Las IDPs exploran superficies de energía libre más planas respecto de las proteínas plegadas, lo cual está asociado a fluctuaciones conformacionales que permiten su función biológica (Csermely et al. 2010). La flexibilidad estructural les permite a las IDPs explotar mecanismos funcionales inaccesibles y complementarios a los de las proteínas plegadas.

El hecho de estar desplegadas les confiere características particulares como:

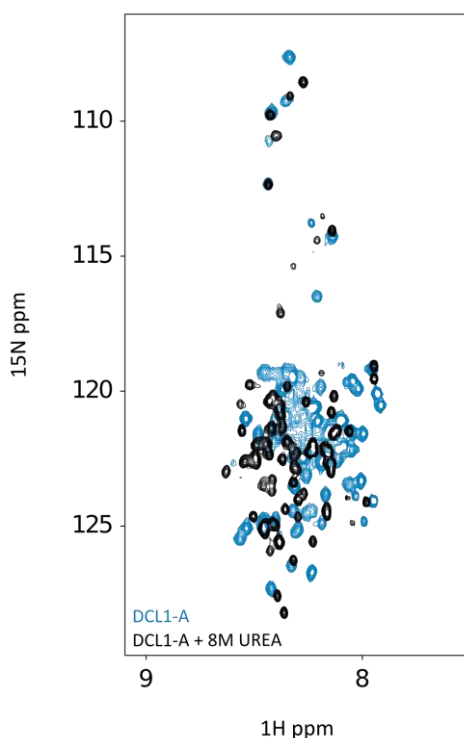
1. Alta velocidad para la formación del complejo con el blanco por poseer un radio de captura mayor al de una proteína plegada (mecanismo de “pesca con mosca”) (Shoemaker et al. 2000). 
2. Alta superficie de interacción por residuo (Uversky 2015).
3. Posibilidad de interaccionar con varias proteínas blanco distintas, lo cual las convierte en muchos casos en el centro de redes regulatorias (Uversky et al. 2005).
4. Posibilidad de plegarse total o parcialmente y de distintas maneras dependiendo del ligando que estén reconociendo (Kriwacki et al. 1996).
5. Disociación entre especificidad y afinidad debida a que a pesar de presentar plegamiento específico según el ligando, muchas de estas proteínas presentan velocidades de disociación rápidas, lo que las hace especialmente adecuadas para procesos regulatorios (Dunker et al. 2001).

Por lo establecido en el capítulo anterior sabemos que DCL1-A se encuentra desplegada en solución, pero desconocemos el grado de desorden real en la cadena. La espectroscopia de resonancia magnética nuclear es la técnica por excelencia para estudiar este tipo de sistemas, en los cuales es necesario que las moléculas estén en solución dado que su comportamiento está determinado por la alta flexibilidad de la cadena. Procedimos entonces a adquirir parámetros de RMN sobre DCL1-A libre para obtener información sobre su estado conformacional a una mayor resolución.



#### 4.1. DCL1-A no está completamente desplegada

Habiendo determinado que el espectro HN-HSQC de DCL1-A corresponde al de un polipéptido desestructurado, adquirimos un espectro HN-HSQC en presencia de Urea 8 M con la intención de evidenciar la existencia de estructura residual. El espectro obtenido es muy diferente al de la forma nativa (Figura 4.1). Este gran cambio nos indica que **la forma nativa libre de la proteína posee un cierto grado de plegamiento** que es desplegado por efecto de la Urea.



**Figura 4.1 DCL1-A no se encuentra completamente desestructurada.**

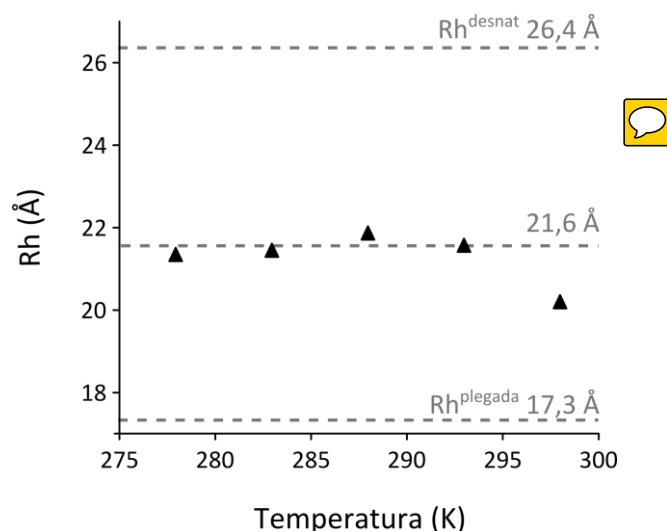
*Espectros HN-HSQC de DCL1-A libre (azul) y en presencia de Urea 8M (negro). La diferencia entre los espectros indica que la forma libre de la proteína en solución presenta una estructura residual.*

#### 4.2. Grado de compactación

Otra forma de estudiar cuan desplegada se encuentra la cadena es determinando el radio hidrodinámico ( $R_h$ ) de la molécula. El  $R_h$  es el radio de una esfera ideal que difundiría en solución a la misma velocidad que la molécula de interés. Por lo tanto, aunque no se trate de la medida real del radio de una proteína no globular, como su difusión está relacionada con su forma no esférica, es muy útil como una medida simple del grado de compactación de proteínas desordenadas. Se sabe que las proteínas desestructuradas presentan un grado de compactación mayor al de proteínas desplegadas químicamente. Incluso se ha demostrado que este grado de compactación depende también de la fracción de residuos de prolina y de la carga neta de la cadena (Marsh & Forman-kay 2010).

Realizamos la determinación del radio hidrodinámico utilizando experimentos de RMN con gradiente de campo pulsado, incorporando dioxano como patrón interno en la muestra (Wilkins et al. 1999). Este patrón interno de radio hidrodinámico conocido nos permite realizar una

medida relativa del radio de la proteína, independizándonos de la viscosidad de la solución. Los valores obtenidos se muestran en la Figura 4.2.



**Figura 4.2** Radio Hidrodinámico de DCL1-A.

Triángulos: valores de  $Rh$  experimentales obtenidos utilizando Dioxano como patrón interno (Wilkins et al. 1999). Líneas grises interrumpidas: valores de referencia teóricos según Marsh y Forman-kay:  $Rh^{desnat}$ ,  $Rh$  para la proteína completamente desnaturalizada (Ecuación 4.2);  $Rh^{plegada}$ ,  $Rh$  para la proteína plegada (Ecuación 4.1); 21,6 Å es el valor calculado para proteínas desestructurados teniendo en cuenta la proporción de residuos de prolina y la carga neta (Ecuación 4.4) (Marsh & Forman-kay 2010).

En el gráfico se muestran los valores experimentales obtenidos a cinco temperaturas diferentes. Los valores de referencia (en gris) corresponden a los valores teóricos calculados según las siguientes relaciones empíricas obtenidas por Marsh y Forman-kay (Marsh & Forman-kay 2010):

4.1

$$Rh^{plegada} = 4,92 N^{0,285}$$

4.2

$$Rh^{desnat} = 2,33 N^{0,549}$$

4.3

$$Rh^{IDP} = 2,49 N^{0,509}$$

4.4

$$Rh^{IDP} = (1,24 P_{Pro} + 0,904)(0,00759 |Q| + 0,963) S_{his} 2,49 N^{0,509}$$

En todas las ecuaciones  $N$  corresponde al número de residuos. La ecuación 4.1 surge del ajuste de la relación empírica entre los datos de  $Rh$  y el número de residuos para proteínas plegadas. La 4.2 se basa en datos de proteínas en condiciones fuertemente desnaturalizantes. La ecuación 4.3 es la relación empírica para proteínas intrínsecamente desordenadas (IDPs) naturales. Marsh y Forman-kay determinaron que mientras la correlación de los valores experimentales con las dos primeras ecuaciones es aceptable, en el caso de las IDPs la ecuación 4.3 no es

suficientemente robusta. En la ecuación 4.4 se intenta corregir el desvío. Para esto se tienen en cuenta contribuciones que dependen de la secuencia de la cadena, particularmente de la fracción de residuos de prolina ( $P_{Pro}$ ) y la carga neta ( $Q$ ).  $S_{His}$  es un factor que depende de si la muestra conserva la cola de Histidina utilizada para la purificación. En nuestro caso la cola es eliminada, por lo tanto  $S_{His}$  tiene un valor de 1.

En la Figura 4.2 observamos que **los valores experimentales de  $R_h$  para DCL1-A varían alrededor de  $21,2 \pm 0,6 \text{ \AA}$** . Por comparación de estos con los valores teóricos calculados según la longitud de la cadena, podemos concluir que **la forma libre de la proteína es más compacta a lo esperado para la forma completamente desplegada ( $26,4 \text{ \AA}$ )**. Por otro lado, **es menos compacta que una proteína plegada de la misma longitud ( $17,3 \text{ \AA}$ )**.

En cuanto a la correlación con el valor calculado para una proteína intrínsecamente desordenada podemos decir que la ecuación de Marsh y Forman-kay que tiene en cuenta factores de la secuencia (Ecuación 4.4) predice el valor con muy buena exactitud. Esta ecuación arroja un valor de  $21,6 \text{ \AA}$ , mientras el valor experimental está en  $21,2 \pm 0,6 \text{ \AA}$ . DCL1-A contiene sólo dos residuos de Prolina y una carga neta de +2, por lo que en principio no esperamos una gran contribución por parte de estos factores. Sin embargo, al calcular el valor según la ecuación 4.3, en la que no se consideran las contribuciones que dependen de la secuencia, obtenemos un valor de  $R_h$  de  $23,6 \text{ \AA}$ . Por lo tanto, **a pesar de ser pequeña, la contribución de las prolinas y la carga neta estaría jugando un rol en el grado de compactación de la molécula**.

La observación de que el dominio no se encuentra completamente desestructurado y presenta un cierto grado de compactación nos da el indicio de que está explorando algún tipo de estructura de forma transitoria. Para probar este supuesto, procedimos a caracterizar por RMN el tipo de conformaciones que explora la proteína a nivel de cada residuo de aminoácido.

### 4.3. DCL1-A explora conformaciones de hélice $\alpha$ en el C-terminal

#### 4.3.1. Asignación del esqueleto carbonado

El primer paso necesario para analizar parámetros de RMN y obtener información específica por residuo, es la asignación de las resonancias correspondientes a cada uno de los átomos involucrados en el enlace peptídico de cada residuo de la proteína.

Para esto produjimos una muestra del DCL1-A marcada con isótopos magnéticamente activos de Nitrógeno y Carbono, expresando la proteína en medio mínimo M9 suplementado con amonio marcado con  $^{15}\text{N}$  y glucosa marcada con  $^{13}\text{C}$ . Para realizar la asignación adquirimos un conjunto de experimentos de triple resonancia estándar que permitieron identificar las resonancias correspondientes al H y N amida, C carbonílico, C alfa, y C beta de cada residuo y sus vecinos. Los experimentos fueron: HNCa, HN(Co)Ca, HNCo, HN(Ca)Co, HNCaCb, HN(Co)CaCb, a 298 K.

Por tratarse de una proteína desestructurada los espectros presentan muchas señales superpuestas, la variación en intensidad de las señales es mayor a lo habitual y los desplazamientos químicos son degenerados (Figura 4.3). Las características mencionadas dificultan el análisis de los espectros y la asignación inequívoca de las señales. Para refinar la asignación inicial y poder discriminar entre sistemas de espín solapados adquirimos dos espectros adicionales. Un espectro HN(Ca)NNH que nos brinda información de correlación de cada par N-H con sus vecinos, y un espectro (H)CC(Co)NH que nos proporciona información de identidad del residuo anterior.



enlaces peptídicos tienen valores característicos según se encuentren en una estructura tipo hélice alfa o lámina beta. Para evaluar el tipo de estructura secundaria en que se encuentra cada residuo se calcula la diferencia entre los desplazamientos obtenidos experimentalmente y los de referencia para proteínas tipo ovillo estadístico (Wishart et al. 1995). En los casos de los carbonos alfa y carbonílicos esta diferencia da un resultado positivo si se trata de una hélice alfa, o negativo si se trata de una lámina beta. Esta misma diferencia es de signo opuesto para los carbonos beta. A la diferencia con los valores de referencia se la denomina desplazamiento químico secundario. En el caso de proteínas desestructuradas o parcialmente plegadas los residuos no adoptan conformaciones rígidas. Para este tipo de proteínas los valores de los desplazamientos químicos secundarios se encuentran cercanos a cero por estar más cercanos al ovillo estadístico. De todas formas, los valores obtenidos para proteínas desestructuradas reportan sobre la tendencia de cada residuo a poblar transitoriamente conformaciones correspondientes a cada tipo de estructura secundaria (Dyson et al. 2004).

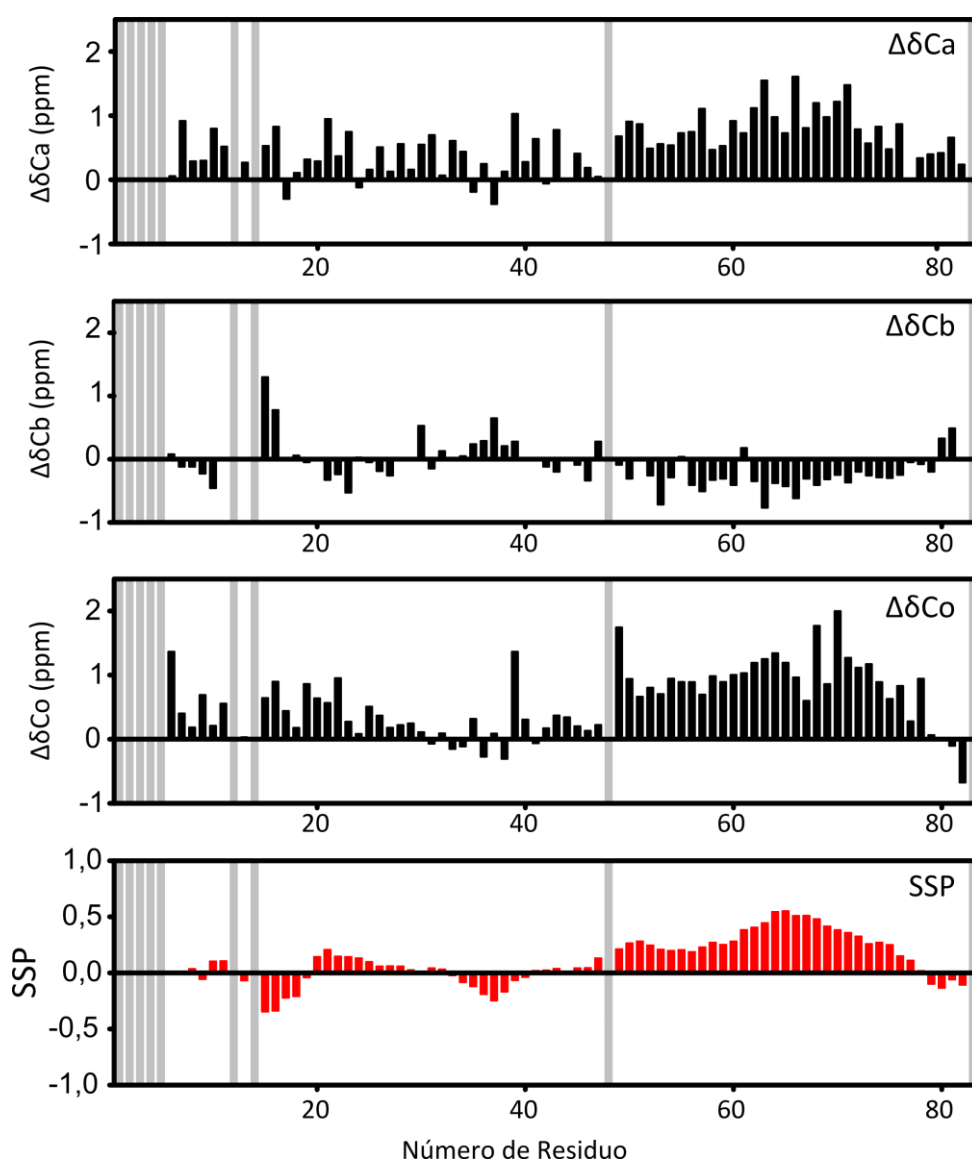


Figura 4.5 DCL1-A libre explora conformaciones de hélice alfa en el extremo C-terminal.

Tres paneles superiores: desplazamientos químicos secundarios para el Carbono alfa, Carbono beta y carbonilo. Panel inferior: probabilidad de poblar estructura secundaria calculada con el programa SSP

(Marsh et al. 2006). *Las barras grises son resonancias no asignadas.*

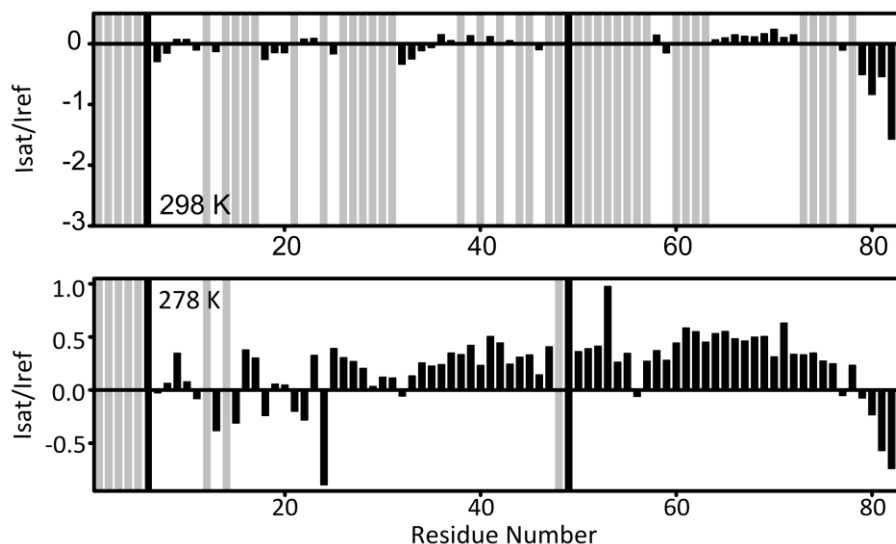
En la Figura 4.5 se muestran los valores de desplazamientos químicos secundarios obtenidos para DCL1-A. Aunque pequeño, el desvío respecto de los valores de referencia para un ovillo estadístico reportan una cierta tendencia a explorar conformaciones de estructura secundaria particulares. Los núcleos de esqueleto carbonado presentan diferente sensibilidad al tipo de estructura secundaria de forma aminoácido específica (Wang & Jardetzky 2002). Para facilitar el análisis de valores de desplazamientos químicos Forman-kay y sus colaboradores desarrollaron el programa SSP (Marsh et al. 2006). SSP calcula la probabilidad de explorar conformaciones de estructura secundaria definidas combinando datos de los desplazamientos químicos de Carbono alfa, beta y carbonílico en un único valor de forma residuo específica. Aplicando este método obtuvimos los valores del panel inferior de la Figura 4.5. Valores de SSP positivos corresponden a estructura tipo hélice y valores negativos corresponden a estructuras extendidas tipo lámina beta. Los valores obtenidos con SSP para DCL1-A indican que **la región comprendida entre los residuos 49 y 78 presenta una tendencia definida a formar hélice alfa a 298 K**. La región N-terminal no presenta una tendencia a adoptar ninguna conformación particular. Para ampliar la descripción de la estructura parcial de la región C-terminal decidimos determinar el grado de rigidez de la cadena.

#### 4.4. Dinámica de la cadena de DCL1-A

Distintos parámetros de RMN contienen información sobre la dinámica de la cadena. Una forma de medir dinámica en el orden de los pico a nanosegundos es a través del Efecto Nuclear Overhauser (NOE) heteronuclear entre el H y N amida del enlace peptídico. La transferencia de magnetización entre los núcleos por NOE depende del tiempo de correlación rotacional de cada par de núcleos. Para estudiar este efecto adquirimos espectros HN-HSQC partiendo de magnetización en <sup>15</sup>N, uno presaturando los protones amida y otro (de referencia) sin presaturación. En el espectro con presaturación, el sistema se encuentra en un estado estacionario en el cual la magnetización de los núcleos de protón es cero. En estas condiciones ocurre la transferencia de magnetización por NOE desde el H al N amida. El NOE entre estos dos núcleos es la relación de intensidades entre las señales del espectro presaturado y sin saturar. Aquellos enlaces H-N que se mueven con una velocidad mayor al tiempo de correlación rotacional de la molécula completa presentan una relación de intensidades menor a la de regiones menos flexibles. Valores entre 0,8 y 1 indican rigidez típica de regiones estructuradas, mientras que valores menores a cero indican alta flexibilidad. Los extremos N y C-terminal de las proteínas plegadas presentan generalmente valores de NOE heteronuclear bajos o negativos (Kleckner & Foster 2011).

Adquirimos los espectros necesarios para calcular el NOE heteronuclear entre el H y el N amida a 298 K. Los valores obtenidos se muestran en el panel superior de la Figura 4.6. Como se puede ver en el gráfico, los valores obtenidos son cercanos a cero a lo largo de la mayor parte del polipéptido, lo que estaría sugiriendo que **la proteína se encuentra desestructurada pero no es del todo flexible**. Los valores negativos obtenidos para el extremo C-terminal son típicos de los extremos altamente flexibles de las cadenas polipeptídicas en general. El espectro presaturado presenta una cantidad de señales mucho menor al de referencia, debido a que los valores de NOE son cercanos a 0. Esta es la razón por la cual muchos valores no pudieron ser determinados (Barras grises en Figura 4.6). Teniendo en cuenta que un descenso en la temperatura podría aumentar la rigidez de la cadena permitiendo la visualización de una mayor cantidad de señales en el espectro presaturado, realizamos el mismo experimento a 278 K. Los

valores obtenidos a menor temperatura se muestran en el panel inferior de la Figura 4.6. Como esperábamos, **la flexibilidad de la cadena disminuye al bajar la temperatura**. Los valores positivos son menores a 0,8, lo cual se corresponde con regiones no estructuradas. Sin embargo, al comparar los valores obtenidos a 278 K con los de 298 K, observamos que desde el residuo 25 al 76 la rigidez de la cadena es mayor a menor temperatura.



**Figura 4.6** La flexibilidad de la cadena de DCL1-A disminuye a 278 K.

*Valores de NOE heteronuclear entre H y N amida en función del número de residuo de DCL1-A a 298 K (panel superior) y 278 K (panel inferior). Barras grises son señales ausentes. Barras negras corresponden a prolina.*

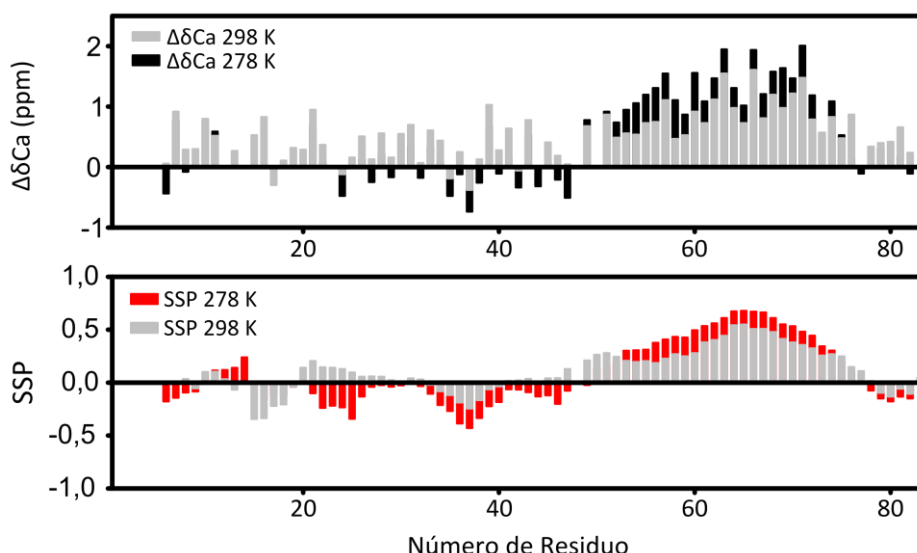
La disminución de la flexibilidad de la cadena podría deberse a la estabilización de la estructura residual que observamos en la sección anterior. Esta idea nos indujo a repetir la asignación de las resonancias del esqueleto carbonado de la proteína a 278 K, con la intención de determinar el tipo de conformación que están explorando los residuos a menor temperatura y comparar con lo obtenido previamente a 298 K.

#### 4.5. Estabilización de la estructura secundaria a 278 K

Empleando la misma metodología que en el apartado 4.3.1 llevamos a cabo la asignación de las resonancias de los núcleos que forman parte del esqueleto peptídico de la proteína, sólo que en esta ocasión los espectros fueron adquiridos a 278 K. Una vez asignadas las resonancias para los núcleos del esqueleto peptídico de la cadena, calculamos los desplazamientos químicos secundarios y la probabilidad de adquirir estructura secundaria con el programa SSP (Marsh et al. 2006). En la Figura 4.7 ilustramos los resultados obtenidos.

Al comparar los valores de los desplazamientos químicos secundarios a 278 K con los obtenidos a 298 K (apartado 4.3.2) podemos observar que, a pesar de la disminución de la temperatura, la región N-terminal no adquiere una estructura secundaria definida. Por otro lado, los valores en la región C-terminal aumentan en comparación con los valores a 298 K (Figura 4.7, panel superior). Este aumento en el valor absoluto de los desplazamientos químicos secundarios se traduce en el aumento de la probabilidad de poblar conformaciones con estructura secundaria

definida según el método de SSP (Figura 4.7, panel inferior). Los resultados indican una estabilización de la hélice alfa en la región C-terminal poblada de forma transitoria a 298 K. La estabilización de la hélice se condice con el aumento de la rigidez según los valores de NOE heteronuclear NH amida obtenidos en la sección 4.4.



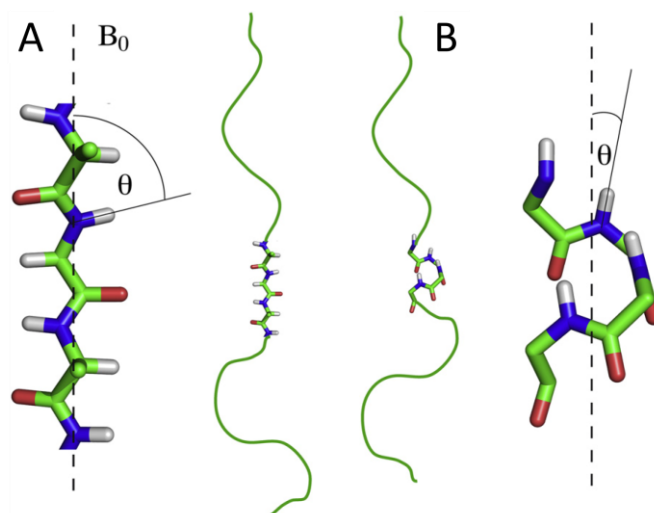
**Figura 4.7** A menor temperatura se estabiliza la hélice alfa en la región C-terminal de DCL1-A.

*Panel superior:* Valores de desplazamientos químicos secundarios de Carbono alfa por residuo de DCL1-A libre a 278 K (negro) y 298 K (gris). *Panel inferior:* probabilidad de poblar conformaciones con estructura secundaria definida calculada con el programa SSP (Marsh et al. 2006). Valores positivos indican conformación de hélice alfa y negativos corresponden a conformaciones extendidas o de lámina beta.

#### 4.6. Orientaciones en el espacio preferidas por DCL1-A

Uno de los métodos de RMN para evidenciar orientaciones preferidas por las distintas regiones de una cadena es la determinación de los acoplamientos dipolares residuales (RDCs). En muestras de proteína en solución acuosa los acoplamientos dipolares se promedian a cero debido a la reorientación isotrópica de las moléculas dentro del tubo. Cuando se coloca la muestra en un medio anisotrópico, las moléculas se alinean parcialmente respecto del campo magnético y es posible medir el acoplamiento dipolar residual. Dado que el acoplamiento depende de la orientación respecto del campo magnético externo, su valor nos brinda información acerca de la orientación relativa de cada enlace respecto del marco de la proteína completa (Bax & Grishaev 2005). Valores de RDCs cercanos a cero o negativos surgen conformaciones extendidas (Louhivuori et al. 2003; Mohana-Borges et al. 2004; Jensen et al. 2009), mientras que valores altos positivos se atribuyen a la presencia de elementos de estructura secundaria o terciaria residual (Bertoncini et al. 2005; Mohana-Borges et al. 2004). Estos valores derivan de los ángulos formados entre el enlace HN-N amida y el campo magnético externo dependiendo del tipo de conformación de la molécula, como se ilustra en la Figura 4.8 (Jensen et al. 2009). Una molécula parcialmente alineada en una conformación extendida experimentará un ángulo  $\Theta$  cercano a  $90^\circ$  (Figura 4.8, A) dando como resultado un valor de RDC cercano a cero o negativo, mientras que una molécula en una conformación tipo hélice presentará un ángulo  $\Theta$  pequeño (Figura 4.8, B), dando como resultado un valor de RDC alto positivo.

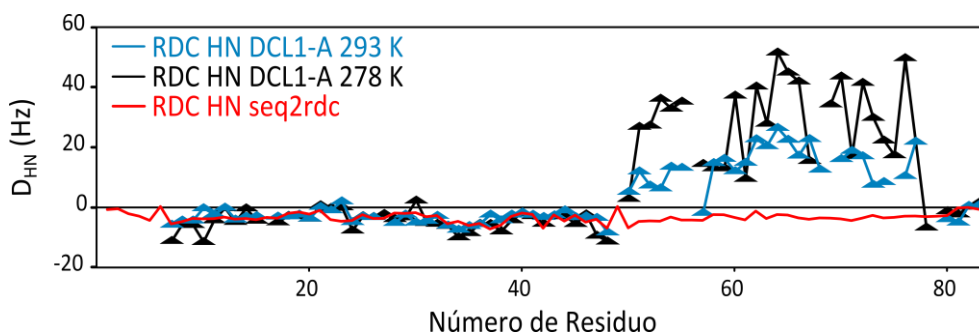




**Figura 4.8** Origen de los valores de RDCs en medio anisotrópico.

A. Angulo ( $\theta$ ) formado entre el par HN-N y el campo magnético externo ( $B_0$ ) para una conformación extendida. B. Angulo formado entre el par HN-N y el campo magnético externo para una conformación tipo hélice (Jensen et al. 2009).

Con la intención de complementar la información brindada por los desplazamientos químicos secundarios, procedimos a medir los RDCs HN-N sobre los grupos amida de la cadena principal en una muestra de DCL1-A alineada parcialmente en fase anisotrópica C8E5-Octanol. En la Figura 4.9 se muestran los valores obtenidos a 293 K y a 278 K.



**Figura 4.9** Comparación de los RDCs de DCL1-A a 293 y 278 K.

Acoplamiento dipolares residuales medidos sobre la proteína parcialmente alineada en fase C8E5-Octanol a 293 K (azul) y 278 K (negro). En rojo se muestran los valores calculados con el algoritmo seq2rdc (Huang et al. 2013).

A primera vista podemos distinguir dos regiones marcadamente diferentes. La región desde el extremo N-terminal hasta el residuo 48 presenta valores de RDCs negativos y cercanos a cero. Estos valores son los esperados para cadenas tipo ovillo estadístico. A partir del residuo 50 se observa claramente que hacia el extremo C-terminal los valores son mayores. Esta región es la que presenta la mayor tendencia a explorar estructuras tipo hélice alfa según los valores de desplazamientos químicos secundarios calculados en apartados anteriores.

En la Figura 4.9 incluimos como referencia los valores correspondientes a los RDCs predichos a partir de la secuencia de la proteína, empleando el algoritmo seq2rdc (Huang et al. 2013). Seq2rdc fue desarrollado para calcular valores teóricos de RDCs para proteínas intrínsecamente

desordenadas, por lo tanto los valores graficados en rojo corresponden a los valores esperados en el caso en que DCL1-A estuviese completamente desestructurada. **Al comparar los valores esperados con los experimentales podemos observar que seq2rdc predice los RDCs de la región N-terminal con exactitud, indicando que la cadena se encuentra desestructurada. Por otro lado, observamos un gran desvío de los valores experimentales respecto de los predichos en la región C-terminal, lo cual pone en evidencia la tendencia a explorar conformaciones más complejas.**

Al analizar la dependencia con la temperatura, observamos que los valores de la región N-terminal hasta el residuo 48 no sufren grandes cambios. Esto se contrasta con la región a partir del residuo 50, que muestra un marcado aumento en los valores de RDC con la disminución de la temperatura. Este fenómeno también se condice con lo observado en el cambio en los desplazamientos químicos secundarios con la disminución de la temperatura. En conjunto, podemos concluir que **a 278 K la tendencia a formar hélice alfa en la región C-terminal se ve aumentada respecto a temperaturas más elevadas, mientras que la región N-terminal permanece desestructurada.**

#### 4.7. DCL1-A libre: un ensamble de múltiples conformaciones en intercambio

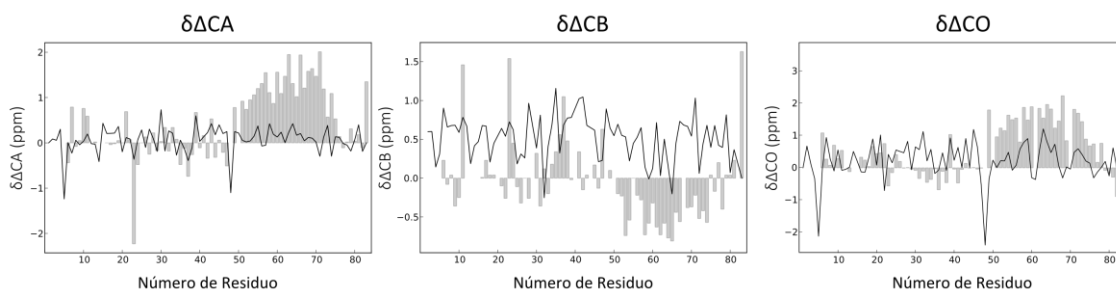
Dada la naturaleza desplegada de DCL1-A no es posible describirla como una única especie de conformación definida. En estos casos la forma de describir el estado de la proteína libre en solución es a través de un conjunto de múltiples conformaciones que se encuentran en intercambio entre sí. Lo que se busca es seleccionar combinaciones de conformaciones que reproduzcan los datos experimentales. De esta manera podemos obtener un modelo que represente de forma satisfactoria el comportamiento conformacional de la proteína en su estado nativo.

##### 4.7.1. Construcción del ensamble conformacional de DCL1-A

Nuestro primer paso fue generar un gran número de conformaciones posibles de la proteína al azar. Para esto construimos un ensamble de 10.000 estructuras utilizando Flexible-Meccano (FM) (Jensen et al. 2008; Bernadó et al. 2005). El algoritmo FM fue especialmente desarrollado para la construcción de ensambles conformacionales de proteínas desestructuradas. FM construye cada conformación seleccionando un par de ángulos diedros  $\Phi/\Psi$  al azar para cada residuo de la proteína en estudio. Los pares de ángulos  $\Phi/\Psi$  son seleccionados según la identidad del residuo a partir de una base de datos de pares de ángulos posibles para cada tipo de residuo. La base de datos consiste en los ángulos diedros de residuos en regiones sin estructura secundaria definida, obtenidos de estructuras cristalográficas depositadas en el PDB. FM utiliza un pseudoátomo centroide para representar el volumen de la cadena lateral de cada residuo. Cada vez que el programa asigna un par de ángulos diedros a un residuo revisa si existen colisiones estéricas entre las cadenas. Si encuentra colisiones estéricas descarta el par de ángulos y selecciona otro de la base de datos. De esta forma sólo conserva los conformeros físicamente posibles. Una vez obtenidos los conformeros al azar con FM reemplazamos el pseudoátomo centroide por las cadenas laterales completas modelándolas con SCCOMP (Eyal et al. 2004).

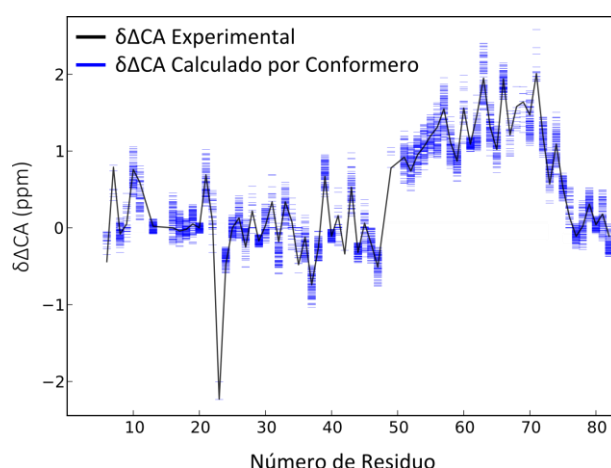
Una vez obtenidas las 10.000 estructuras procedimos a calcular los desplazamientos químicos teóricos para los átomos del esqueleto peptídico empleando el programa de predicción de desplazamientos químicos SPARTA+ (Shen & Bax 2010). Obtenidos los desplazamientos químicos para los 10.000 conformeros, calculamos el promedio de los valores correspondientes

al ensamble completo y contrastamos con los valores experimentales (Figura 4.10). De la diferencia observada entre los valores predichos con los valores experimentales podemos concluir que **este conjunto de estructuras no representa correctamente el estado de la proteína en solución**. Estos resultados confirman nuestras observaciones previas que indican que la forma libre de la proteína no se encuentra completamente desestructurada.



**Figura 4.10** Un ensamble de estructuras al azar no describe correctamente el estado conformacional de DCL1-A libre.

Las barras grises corresponden a los valores de desplazamientos químicos secundarios determinados experimentalmente a 278 K en la sección 4.5 para la proteína libre. Las curvas negras corresponden al promedio de los valores de desplazamientos químicos secundarios predichos para los confórmers del ensamble de 10.000 estructuras al azar generadas con FM.

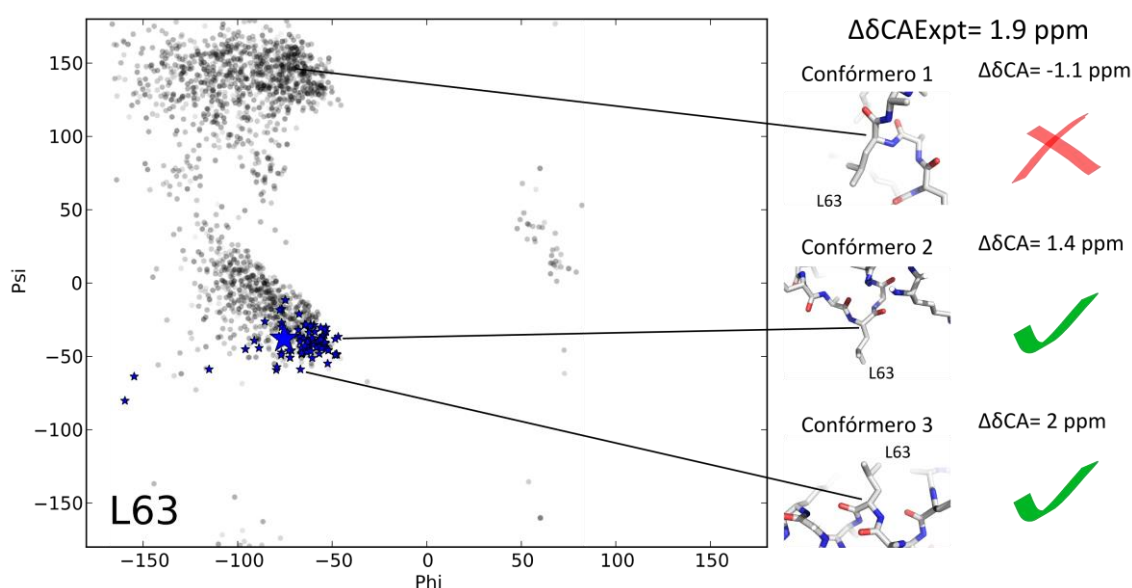


**Figura 4.11** Confórmers del ensamble de FM que reproducen los datos experimentales

En negro se grafican los valores de desplazamientos químicos secundarios de carbono alfa experimentales a 278 K. Cada línea azul corresponde al valor de desplazamiento químico secundario calculado para cada residuo de un conformero aislado del ensamble.

Dado el origen de la base de datos utilizada por FM para asignar los valores de ángulos diedros, los cónformers en este ensamble exploran una amplia región del diagrama de Ramachandran. Los desplazamientos químicos de los átomos de carbono del esqueleto peptídico dependen principalmente de los ángulos diedros  $\Phi/\Psi$ . Por lo tanto, es posible encontrar entre las estructuras calculadas las conformaciones que reproducen los valores de los desplazamientos químicos experimentales. En la Figura 4.11 se muestran los valores de desplazamientos químicos calculados para cada residuo por confórmers aislados que presentan un desvío menor a 0,5 ppm respecto del valor experimental.

Con el objetivo de generar un nuevo ensamble que represente el estado nativo de DCL1-A, construimos una nueva base de datos de ángulos diedros específica para esta proteína. Para lograr esto, tomamos los pares de ángulos  $\Phi/\Psi$  que corresponden a las conformaciones que mejor reproducen los datos experimentales. En la Figura 4.12 ilustramos el método de selección. Tomamos como máximo 100 valores de  $\Phi/\Psi$  por residuo de confórmeros cuyo desplazamiento químico se desviara **menos de 0.5 ppm** respecto de los valores experimentales. En esta base de datos cada residuo de la proteína tiene asignados un grupo de pares de ángulos específicos. Por ejemplo los ángulos posibles para la leucina 63 no son los mismos que para la leucina 10. Esta característica la hace esencialmente diferente a la base de datos empleada por FM, en la cual los ángulos son específicos para cada tipo de residuo sin diferenciarlos por su posición en la secuencia.

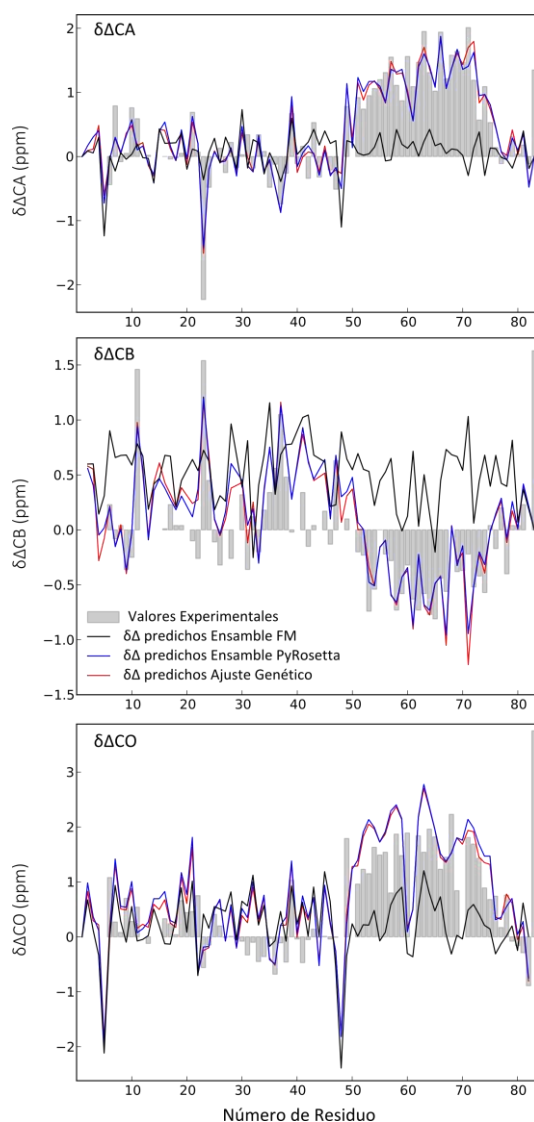


**Figura 4.12 Selección de ángulos diedros para la construcción de una base de datos específica para DCL1-A.**

*En este ejemplo se muestra el diagrama de Ramachandran para la Leucina 63 en el ensamble de conformaciones al azar construido con FM. Las estrellas azules corresponden a los confórmeros que presentan desplazamientos químicos calculados más cercanos a los experimentales. En este ejemplo, los ángulos de la Leucina 63 del confórmero 1 no serían incorporados a la nueva base de datos, mientras que sí lo serían los de los confórmeros 2 y 3.*

Utilizando esta base de datos específica para DCL1-A, procedimos a generar un nuevo ensamble de 10.000 confórmeros empleando PyRosetta (Chaudhury et al. 2010). PyRosetta es una interfase para el software de modelado molecular Rosetta basada en el lenguaje de programación Python. PyRosetta permite escribir rutinas personalizadas de predicción de estructuras y diseñar algoritmos usando las principales funciones de muestreo y medición de energía de Rosetta. Para construir el nuevo ensamble programamos la rutina de PyRosetta adjunta en el Apéndice I. Brevemente, en esta rutina cada confórmero es generado con Rosetta atribuyendo a cada residuo un par de ángulos diedros al azar a partir de la base de datos específica que construimos a partir del primer ensamble de confórmeros al azar. Luego se realiza una optimización de la conformación cambiando ángulos al azar por otros de la misma base de datos y comparando la energía del nuevo confórmero con el anterior. Se conserva el cambio sólo si la energía disminuye. Pudimos determinar que realizadas 200 vueltas de optimización la energía de los confórmeros es estable. La función de energía utilizada incluye términos correspondientes a puentes de Hidrógeno, interacciones de van der Waals atractivas y

repulsivas, energía de solvatación e interacciones electrostáticas (Chaudhury et al. 2010). Luego de aplicar este método para calcular 10.000 nuevas estructuras, calculamos los desplazamientos químicos teóricos promedio para el ensamble completo usando SPARTA+. En la Figura 4.13 mostramos los valores obtenidos (curva azul). **Por contraste con los valores obtenidos para el ensamble al azar de FM (curva negra) podemos concluir que el nuevo ensamble representa el estado de la proteína nativa con mayor fidelidad.**



**Figura 4.13** Los valores de desplazamiento químico secundario promedio para el ensamble construido por PyRosetta reproducen los valores experimentales con mayor fidelidad que los del ensamble de conformeros al azar construido con FM.

*Panel superior: desplazamientos químicos secundarios de Carbono alfa; panel medio: desplazamientos químicos secundarios de Carbono beta; panel inferior: desplazamientos químicos secundarios de Carbono carbonílico. Barras grises: valores experimentales a 278 K. Curva negra: valores promedio para el ensamble construido con FM. Curva azul: valores promedio para el ensamble construido con PyRosetta. Curva roja: valores promedio para el cromosoma de 100 conformeros que mejor ajusta a los valores experimentales como resultado de un algoritmo genético.*

#### 4.7.2. Determinación de subensambles de confórmeros seleccionados por un algoritmo genético

Los valores de desplazamiento químico son producto del promedio de estructuras presentes en la solución y, a su vez, estas estructuras pueden estar presentes en distintas proporciones en el equilibrio. Decidimos buscar entonces una forma de ajustar la representación de los distintos confórmeros dentro del ensamble con la intención de mejorar el ajuste de los valores calculados a los valores experimentales. Para esto empleamos un algoritmo genético que nos permitiera seleccionar conjuntos de confórmeros que en promedio se acercaran a los valores experimentales. Desarrollamos el algoritmo adjunto en el Apéndice II, en código de Python.

Los subensambles se representan como “cromosomas” cuyos “genes” son los distintos confórmeros. De esta forma podemos definir que tan bien representa cada cromosoma al ensamble real, estimando la desviación de los parámetros calculados para el mismo (desplazamientos químicos) respecto de los valores experimentales. La desviación representa una medida de la “aptitud” de cada cromosoma en el “ambiente” y permite seleccionar los mejores cromosomas, que son los que mejor representan los resultados experimentales.

En nuestro caso generamos 1000 cromosomas de 100 genes (confórmeros) cada uno. Dentro de esta población cada confórmero está repetido 10 veces, para aumentar la probabilidad de que una estructura sea seleccionada más de una vez. Los genes que componen a cada cromosoma son inicialmente seleccionados al azar dentro del conjunto de confórmeros del ensamble.

La selección procede de la siguiente manera. Para cada cromosoma calculamos los valores promedio de desplazamientos químicos y el error respecto de los datos experimentales (“aptitud”). Luego, ordenamos la lista de cromosomas según su aptitud, de manera que los cromosomas que mejor reproducen los valores experimentales quedan en la parte superior de la lista. Una vez ordenados, intercambiamos las posiciones de manera aleatoria entre bloques de diez cromosomas, es decir intercambiamos posiciones entre los cromosomas en las posiciones 1 a 10, entre las posiciones 11 a 20, etc. Este paso permite aumentar la probabilidad de recuperar estructuras “buenas” que puedan haber quedado en cromosomas “malos” en la generación siguiente. Una vez mezclados los cromosomas, se produce el entrecruzamiento en puntos al azar entre cada par de cromosomas dentro de la lista. Es decir, el cromosoma en la posición 1 se entrecruza con el cromosoma en la posición 2, el de la posición 3 con el de la 4, etc. Este proceso se repite, produciendo en cada generación de mezcla y entrecruzamiento una nueva generación de cromosomas.

Aplicamos el algoritmo durante 1000 generaciones. La cantidad de generaciones la definimos luego de realizar varias pruebas y observar que el ajuste no mejoraba notoriamente pasada esta cantidad. En la Figura 4.14 mostramos la evolución del ajuste según pasan las generaciones. Realizamos 5 evoluciones independientes.

La curva roja en la Figura 4.13 corresponde al promedio de los desplazamientos químicos secundarios calculados para el mejor cromosoma obtenido en uno de los ajustes genéticos. Como es evidente, el ajuste no mejora notablemente respecto de los valores promedio obtenidos para el ensamble completo. Entre los distintos ajustes independientes observamos estructuras que se repiten, es decir, son seleccionadas con mayor frecuencia que otras. No observamos una gran repetición de estructuras dentro de un mismo cromosoma. Este resultado

indica que no existen subensambles significativamente más representativos respecto de los datos experimentales dentro del ensamble original.

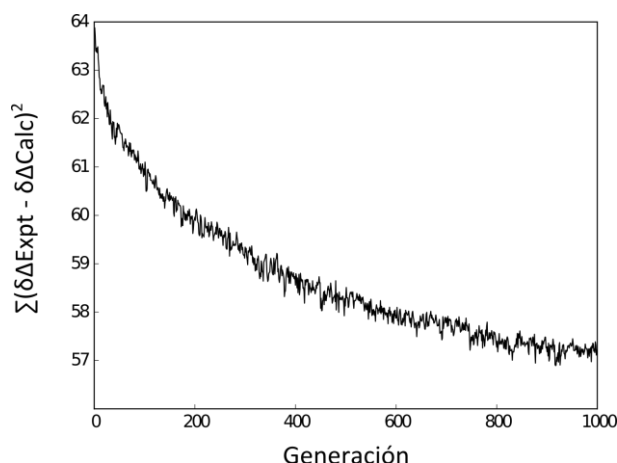


Figura 4.14 Evolución del ajuste por algoritmo genético a través de 1000 generaciones.

#### 4.7.3. Validación cruzada de la calidad del ensamble conformacional

Para obtener una idea de la calidad del ensamble buscamos calcular un parámetro que no hubiéramos empleado como referencia para la selección de los ángulos diedros para construir el ensamble con PyRosetta. Como contábamos con los valores experimentales de RDCs del par HN amida del enlace peptídico a 278 K obtenidos en la sección 4.6, procedimos a calcular este mismo parámetro para cada conformero dentro del ensamble. Realizamos el cálculo empleando el programa PALES (Zweckstetter & Bax 2000). PALES predice la orientación de una molécula en un medio anisotrópico a partir de su estructura. Esto permite calcular los valores esperados de RDCs para dicha estructura. Obtenidos los valores para las 10.000 estructuras en el ensamble, calculamos el promedio y los contrastamos con los valores experimentales, como se muestra en la Figura 4.15.

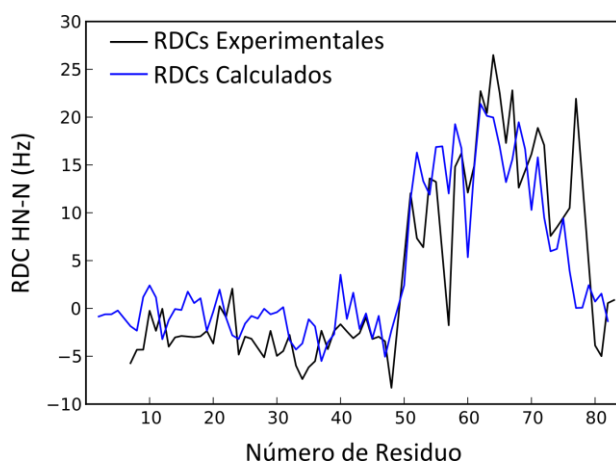


Figura 4.15 Los RDCs calculados para el ensamble de PyRosetta reproducen las características globales de los experimentales.

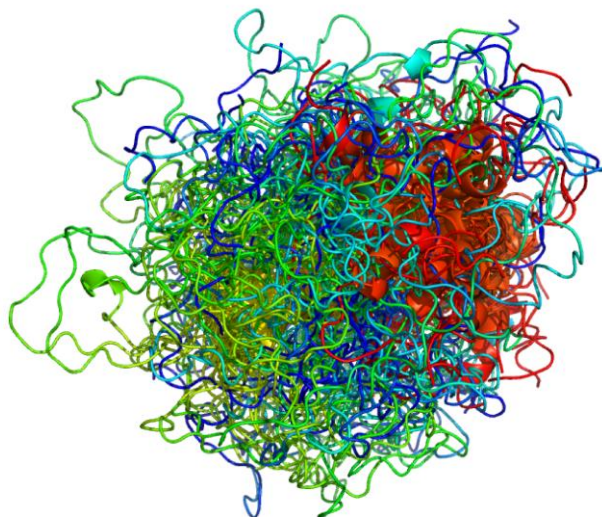
*Azul: RDCs HN-N promedio calculados con PALES (Zweckstetter & Bax 2000) para los 10.000 conformeros del ensamble de PyRosetta. Negro: RDCs HN-N obtenidos experimentalmente a 278 K (sección 4.6).*

La correcta reproducción de los valores experimentales de RDC a partir del ensamble calculado utilizando como restricción los desplazamientos químicos indican que el conjunto de conformeros

generado representa correctamente las características globales del estado nativo de DCL1-A, como también la estructura fina en algunas regiones.

#### 4.7.4. Descripción estructural del ensamble conformacional de DCL1-A

En la Figura 4.16 mostramos la superposición de 100 conformeros, correspondientes al mejor cromosoma producto del ajuste genético del ensamble generado con PyRosetta. Se destaca la naturaleza mayormente desestructurada de la proteína.



**Figura 4.16** Estructuras del ensamble conformacional de DCL1-A generado con PyRosetta.

*Estructuras coloreadas en dirección N-terminal al C-terminal desde azul a rojo, pasando por verde y amarillo.*

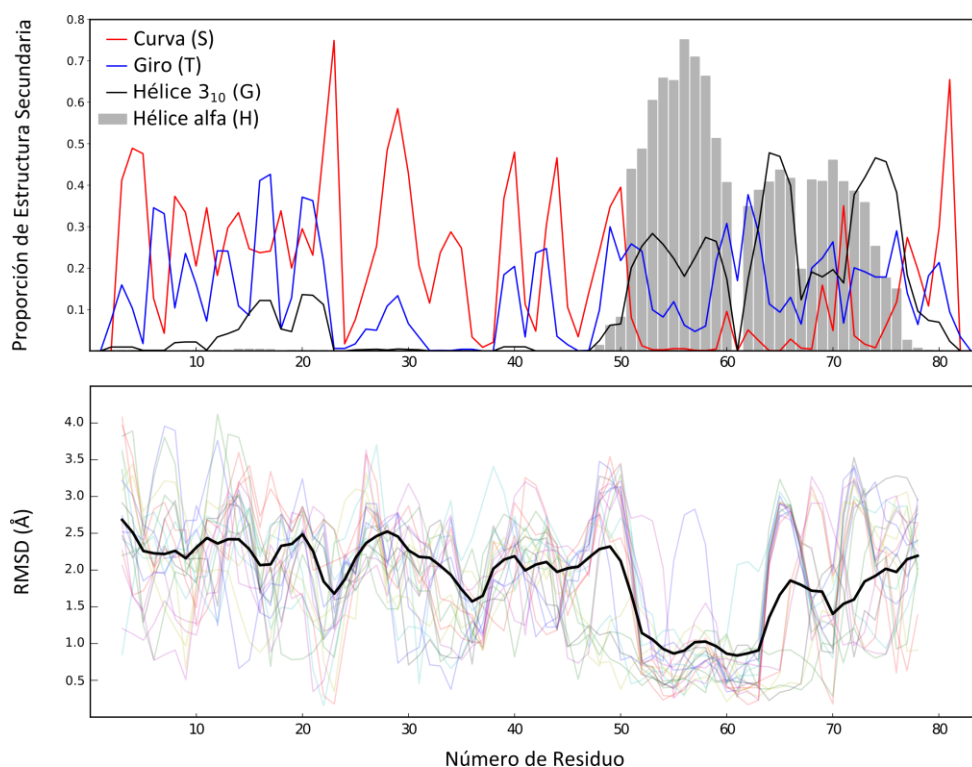
Para describir el ensamble en términos de estructura secundaria empleamos el programa DSSP (Kabsch & Sander 1983), que es el método estándar de asignación de tipo de estructura secundaria a partir de estructuras en tres dimensiones en el PDB. En el panel superior de la Figura 4.17 graficamos las proporciones de tipos de estructura secundaria obtenidas para las 10.000 estructuras del ensamble conformacional de DCL1-A. **Destacamos la dominancia de estructuras tipo bucle (giros y curvas) en la región N-terminal, como esperamos para una cadena flexible como la de DCL1-A. En la región C-terminal observamos que el segmento comprendido entre los residuos 51 a 60 presentan mayoritariamente una conformación tipo hélice alfa. La región entre los residuos 62 a 66 presenta una proporción dominante de estructuras tipo hélice, pero en este caso distribuida entre conformación tipo hélice alfa y hélice  $3_{10}$ . Finalmente observamos dos fragmentos con menor carácter de hélice, desde el residuo 68 al 74 con estructura tipo hélice alfa y desde el residuo 72 al 77 con estructura tipo hélice  $3_{10}$ .**

Con la intención de definir el grado de similitud entre las estructuras en las distintas regiones realizamos un cálculo de RMSD entre las posiciones de los carbonos alfa por residuo dentro de una ventana de alineamiento entre las estructuras de 7 aminoácidos de longitud. Por cuestiones de capacidad de cómputo realizamos este cálculo sobre los 100 conformeros seleccionados por ajuste genético en el apartado 4.7.2. En el panel inferior de la Figura 4.16 mostramos el resultado.

Podemos observar que en la región comprendida entre los residuos 52 a 63 el valor de RMSD es más bajo que en el resto de la proteína. Esto nos estaría indicando que **la región con estructura**



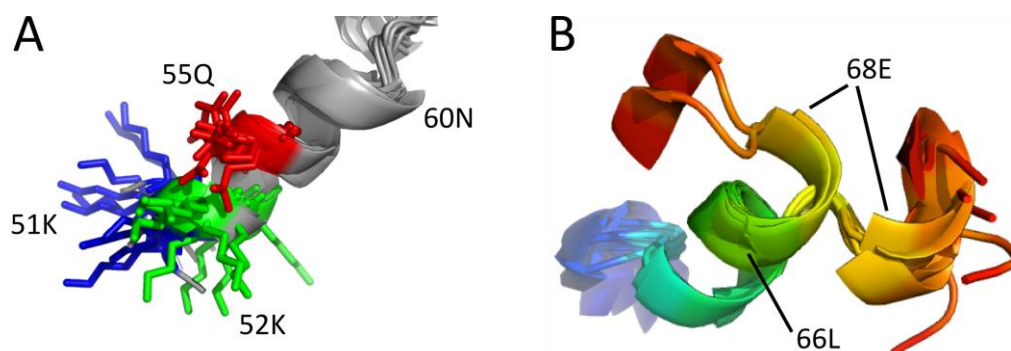
dominante tipo hélice alfa determinada por DSSP se encuentra conservada entre los confórmers. Esta región resulta de interés ya que involucra a las Lisinas 51 y 52 y a Glutamina 55, tres de los residuos putativos de unión al sustrato por dominios tipo dsRBD. En la Figura 4.18 A se muestra el alineamiento estructural de los confórmers entre los residuos 52 a 63. El motivo putativo de unión al sustrato está conformado por las Lisinas 51 (azul), 52 (verde) y la Glutamina 55 (rojo). En dominios de la misma familia este motivo de unión al sustrato actúa como una regla molecular, en la que los residuos homólogos a la Lisina 51 contactan una de las hebras del ARNdh mientras que los homólogos a la Lisina 52 y Glutamina 55 contactan la hebra complementaria, cubriendo el ancho completo del surco mayor de la doble hélice (Masliah et al. 2013). En el ensamble de DCL1-A, las distintas orientaciones de las cadenas laterales de estos residuos en los diferentes confórmers podrían estar confiriendo a la proteína mayor versatilidad al momento de contactar al ARNdh.



**Figura 4.17** Análisis estructural del ensamble conformacional de DCL1-A.

*Panel superior: proporción de tipos de estructura secundaria determinados por DSSP (Kabsch & Sander 1983) para los 10.000 confórmers del ensamble de DCL1-A. Panel inferior: Curva negra: RMSD promedio entre los 100 confórmers dentro del mejor cromosoma seleccionado por ajuste genético del ensamble conformacional de DCL1-A. Curvas de colores: RMSD correspondientes a pares de confórmers por separado. El RMSD fue calculado por residuo dentro de una ventana de alineamiento de 7 residuos de longitud.*

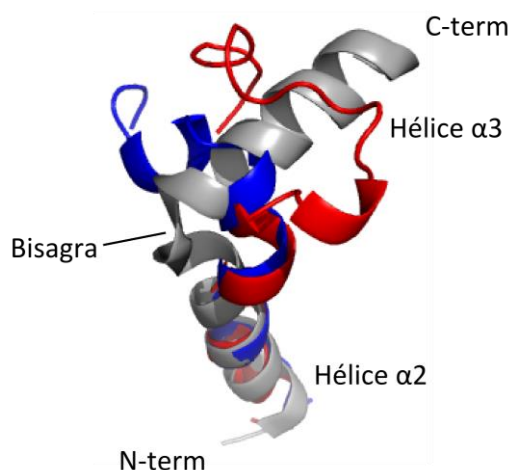
Otra región que se destaca del cálculo de RMSD es la que se encuentra entre los residuos 64 a 67, donde observamos que los RMSD de los confórmers por separado se agrupan en lugar de presentar una distribución heterogénea alrededor del promedio. **Al alinear las estructuras según el segmento comprendido entre los residuos 62 a 67 observamos cómo las estructuras se separan en dos grupos según la dirección que toma la cadena** (Figura 4.18, B). Esta diferencia depende de si la hélice no se interrumpe hasta el residuo 68 o si la hélice se interrumpe en el residuo 66 y la Lisina 67 forma un giro.



**Figura 4.18** Detalles estructurales comunes entre conformeros del ensamble de DCL1-A.

A. Alineamiento estructural del segmento entre los residuos 52 a 63. Se destacan las cadenas laterales de los residuos pertenecientes al motivo putativo de unión al ARNdh: Lisina 51 (azul), Lisina 52 (verde) y Glutamina 55 (rojo). B. Alineamiento estructural del segmento entre los residuos 62 a 67. Se destacan las dos direcciones diferentes que toman las cadenas de los conformeros según la hélice alfa termina en la Leucina 66 o en el Aspártico 68.

Al cambio de dirección de la cadena a la altura de la Leucina 66 lo podemos relacionar con las estructuras conocidas de dos dsRBDs de levadura, el de la ARNasa III Rnt1p (Leulliot et al. 2004) y el de la enzima tipo Dicer Dcr1 (Barraud et al. 2011). Ambos dsRBDs presentan una tercer hélice en el C-terminal, adicional al plegamiento canónico de los dsRBDs. En el dsRBD de Dcr1 esta hélice está involucrada en un sitio de unión a Zinc. No esperamos que encontrar esta función en DCL1-A debido a que en su secuencia no se encuentran ninguna de las tres Cisteínas ni la Histidina que conforman el sitio de unión al metal de Dcr1.



**Figura 4.19** El cambio dirección en la cadena de DCL1-A puede relacionarse con la región bisagra de Rnt1p.

Alineamiento estructural de la región de unión al ARNdh en el extremo N-terminal de la hélice 2 del dsRBD de Rnt1p (gris, 1T4N.pdb) con dos conformeros de DCL1-A representativos de los presentan un quiebre en la hélice en el residuo 66 (rojo) y de los que poseen una hélice continua hasta el residuo 68 (azul).

En la Figura 4.19 mostramos el alineamiento estructural de la región bisagra del dsRBD de Rnt1p con dos conformeros de DCL1-A. En el caso de Rnt1p, la delección de la hélice 3 del dsRBD disminuye la afinidad de la enzima por el sustrato. Esta hélice establece una serie de contactos

con la hélice 1 del dominio, fundamentales para la estabilidad de esta última. Debido a que una de las regiones de reconocimiento del sustrato se encuentra en la hélice 1, los autores atribuyen a la hélice 3 la función de inducir en la hélice 1 la conformación adecuada que le confiere su actividad. En el caso de DCL1-A la hélice 1 está completamente desestructurada mientras que la hélice 3 es mucho más corta que en Rnt1p. Sin embargo, sabiendo que el dominio se pliega en presencia del sustrato, podemos pensar que quizás en la forma plegada final se conserve esta característica estructural. A esto podemos agregar que **el hecho de que las estructuras del ensamble se dividan en dos clases dependiendo de la dirección de la cadena pueda estar relacionado con diferentes maneras de entrar en contacto con el sustrato para que se produzca el plegamiento.**

#### 4.8. Discusión

En esta parte del trabajo nos enfocamos en la descripción detallada del estado conformacional de DCL1-A en su forma nativa libre. Encontramos que la cadena se encuentra predominantemente desordenada y presenta una tendencia a explorar conformaciones helicoidales en la región C-terminal. La región helicoidal involucra a uno de los sitios de unión al ARNdh putativos para la familia de dominios tipo dsRBD, a la cual pertenece la proteína por homología de secuencia. Estas características estructurales podrían ser esenciales para su funcionamiento en el contexto de la enzima DCL1 completa así como también aportan a la discusión general sobre el comportamiento de proteínas desordenadas y su plegamiento.

En cuanto a la unión al sustrato, la naturaleza desordenada de DCL1-A podría estar implicada en la velocidad con que la proteína encuentra al ARNdh en las intermediaciones de la molécula. A esta particularidad de las proteínas desordenadas se la conoce como modelo de unión tipo “pesca con mosca” (del inglés *fly-casting model*). El modelo de unión tipo “pesca con mosca” fue propuesto en el año 2000 por el grupo de Wolynes (Shoemaker et al. 2000). En este modelo Wolynes y sus colaboradores propusieron que una proteína desestructurada posee un radio de captura mayor para un sitio de unión respecto al estado plegado de la misma. En este escenario de unión, el estado desplegado se une débilmente a una distancia relativamente grande, seguido del plegamiento a medida que la proteína se aproxima al sitio de unión. La posibilidad de contactar a su molécula blanco a mayor distancia aumenta las probabilidades de que se forme el complejo. Teniendo en cuenta este modelo, sabemos que dos de los sitios putativos de unión al ARNdh se encuentran en regiones altamente desestructuradas de la proteína, lo que les estaría permitiendo explorar el espacio circundante con un mayor grado de libertad. El tercer sitio de unión se encuentra en una región parcialmente plegada. En los dsRBDs canónicos este sitio es justamente uno de los determinantes que permite unir específicamente a la estructura del ARNdh por sobre otros ácidos nucleicos (Masliah et al. 2013; Vukovic et al. 2014). Las diferentes orientaciones de las cadenas laterales de los residuos pertenecientes a este sitio de unión en DCL1-A podrían estar confiriendo a la vez versatilidad en la unión al sustrato como también un cierto grado de estructura previamente formada para garantizar el reconocimiento.

Por otro lado, la estructura secundaria poblada en la región C-terminal podría estar implicada en la regulación del mecanismo de unión al sustrato. Se conocen varios ejemplos en los que la perturbación de los elementos de estructura secundaria transitoriamente explorados por proteínas desestructuradas afecta su mecanismo de unión y plegamiento (Wright & Dyson 2009; Iešmantavičius et al. 2014). Uno de los casos es la interacción entre el dominio quinasa inducible del factor de transcripción CREB (pKID) con el coactivador CBP. En este sistema el dominio pKID libre es una proteína desestructurada que se pliega al interaccionar con su

dominio blanco (KIX) de CBP (Sugase et al. 2007). Esta interacción ocurre a través de un complejo intermediario desestructurado que fue caracterizado por simulación computacional (Turjanski et al. 2008). La forma libre de pKID cuenta con una región que explora transitoriamente conformaciones tipo hélice alfa. Tanto mutantes que desestabilizan la conformación de hélice de esta región como modificaciones que la estabilizan, disminuyen la afinidad del dominio por KIX (Turjanski et al. 2008). Los autores explican estos fenómenos como cuestiones de la cinética de la formación del complejo. Mientras que las mutantes desestabilizantes aumentan la  $k_{off}$  de la unión, las estabilizantes disminuyen la  $k_{on}$ . Estos resultados nos indican que la dinámica con que las proteínas desestructuradas exploran conformaciones de estructura secundaria se debe a una optimización del evento de unión al blanco. Por lo tanto, podemos pensar que la proporción de estructura secundaria que encontramos en el extremo C-terminal de DCL1-A es probablemente de importancia para la regulación del evento de interacción con el sustrato. Sobre estos interrogantes necesitaríamos avanzar con nuestra caracterización realizando experimentos de cinética de unión al sustrato para la proteína silvestre y mutantes que afecten el grado de estabilidad de la región helicoidal.

## 5. CAPITULO 5: DCL1-A en complejo con el sustrato, mecanismo de unión y plegamiento

Habiendo logrado una descripción detallada de DCL1-A libre, procedimos a la caracterización de la forma unida al sustrato. Como ya demostramos en el capítulo 3, DCL1-A adquiere un plegamiento definido en presencia de ARNdh. Estudiando este mismo complejo empleando RMN logramos obtener información estructural a mayor resolución. Presentamos los datos obtenidos a lo largo de este capítulo.

### 5.1. DCL1-A adquiere un plegamiento tipo dsRBD en presencia del sustrato

#### 5.1.1. Asignación del esqueleto peptídico de DCL1-A en complejo con ARNdh

Para realizar estudios sobre la estructura del dominio en complejo con ARNdh por RMN es necesario asignar las resonancias observadas a los distintos núcleos de la proteína. Al emplear las técnicas estándar de RMN es necesario definir un sistema con el menor peso molecular posible, ya que a mayor tamaño aumenta la velocidad de relajación de los núcleos. El incremento en la relajación resulta en señales cada vez más anchas, disminuyendo su intensidad. Por esta razón decidimos construir el complejo del dominio aislado con el fragmento de ARNdh correspondiente a la región del tallo inferior del precursor del miR172a (pri-miR172a-ti), caracterizado previamente en las secciones 3.3 y 3.4. Este complejo presenta un tamaño final de *ca.* 20 KDa, cercano al límite superior con el que podemos trabajar.

El aumento en la velocidad de relajación de los núcleos se debe principalmente a la interacción dipolar con los núcleos de  $^1\text{H}$ . Una estrategia para reducir la relajación de los núcleos por interacción con los  $^1\text{H}$  alifáticos y aromáticos es utilizar una muestra perdeuterada de la proteína, en la cual todos los núcleos de  $^1\text{H}$  no intercambiabiles con el solvente son reemplazados por núcleos de  $^2\text{H}$ . Para obtener esta muestra realizamos la expresión de la proteína en medio de cultivo mínimo M9 empleando agua, glucosa y amonio deuterados, además de marcados con  $^{15}\text{N}$  y  $^{13}\text{C}$ . Una vez obtenida la muestra de proteína perdeuterada, agregamos 1.5 equivalentes de ARNdh pri-miR172a-ti sin marca isotópica. Luego procedimos a la adquisición de los espectros estándar para asignación del esqueleto carbonado del polipéptido (HN-TROSY, tr-HN(Ca), tr-HN(Co)Ca, tr-HNCo, tr-HN(Ca)Co, tr-HN(Ca)Cb, tr-HN(Co)CaCb).

En la Figura 5.1 se muestra el espectro HN-TROSY de DCL1-A en complejo con el sustrato y el detalle de las resonancias asignadas por residuo. En comparación al espectro de la proteína libre (apartado 4.3.1), se observa la aparición de un nuevo conjunto de señales que presentan una mayor dispersión en ambas dimensiones. **La dispersión de los desplazamientos químicos es la esperada para una proteína plegada.** No se observan picos duplicados para un mismo residuo de la cadena plegada y las señales son finas, sugiriendo que el complejo formado es homogéneo.

Luego del análisis de los espectros conseguimos asignar el 75% de las resonancias esperadas correspondientes al esqueleto peptídico de la forma plegada (Figura 5.1).

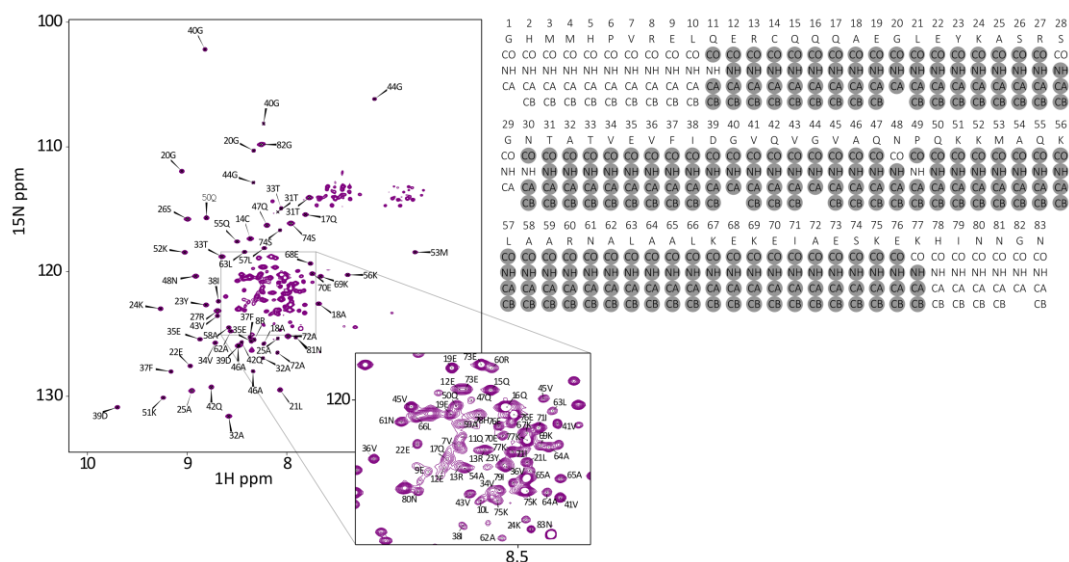


Figura 5.1 El espectro HN-TROSY de DCL1-A en complejo con el sustrato corresponde al de una proteína plegada.

Espectro HN-TROSY de DCL1-A en presencia de 1.5 equivalentes de pri-miR172a-ti. A la derecha se muestra un detalle de las resonancias asignadas (círculos grises).

### 5.1.2. Determinación de la estructura secundaria adquirida

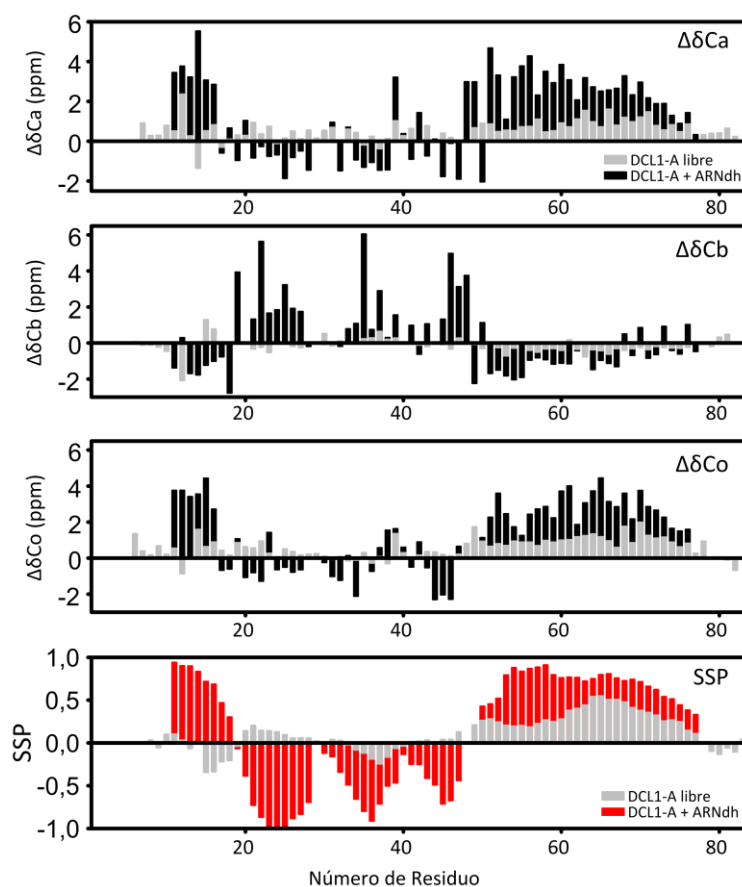


Figura 5.2 DCL1-A en complejo con ARNdH presenta un plegamiento  $\alpha\beta\beta\beta\alpha$ .

*Tres paneles superiores: desplazamientos químicos secundarios para el Carbono alfa, Carbono beta y Carbono carbonílico. Barras negras: DCL1-A + 1.5 equivalentes pri-miR172a-ti, barras grises: DCL1-A libre a 298 K. Panel inferior: probabilidad de poblar estructura secundaria calculada con el programa SSP (Marsh et al. 2006). Barras rojas: DCL1-A + 1.5 equivalentes pri-miR172a-ti, barras grises: DCL1-A libre a 298 K. Valores positivos de SSP corresponden a conformaciones de hélice, valores negativos corresponden a conformaciones extendidas o tipo lámina beta.*

Como describimos anteriormente, los desplazamientos químicos de los carbonos alfa y carbonilo de los enlaces peptídicos tienen valores característicos según se encuentren en una estructura tipo hélice alfa o lámina beta.

**Los valores de desplazamientos químicos secundarios obtenidos para DCL1-A en presencia del sustrato revelan un plegamiento tipo  $\alpha\beta\beta\alpha$ , típico de dominios dsRBD** (Figura 5.2, barras negras). Es notable la diferencia en magnitud con los resultados obtenidos en la sección 4.3 para el dominio en su forma libre desestructurada (Figura 5.2, barras grises). La mayor magnitud indica que los residuos se encuentran formando una estructura estable. De la misma manera que para la forma libre, calculamos la proporción de estructura secundaria para cada residuo con el método de SSP (Marsh et al. 2006). Los valores se muestran en el gráfico inferior de la Figura 5.2 (rojo) en comparación con los valores obtenidos para la forma libre a 298 K (gris). SSP integra la información brindada por los desplazamientos químicos secundarios mostrando claramente el plegamiento  $\alpha\beta\beta\alpha$ .

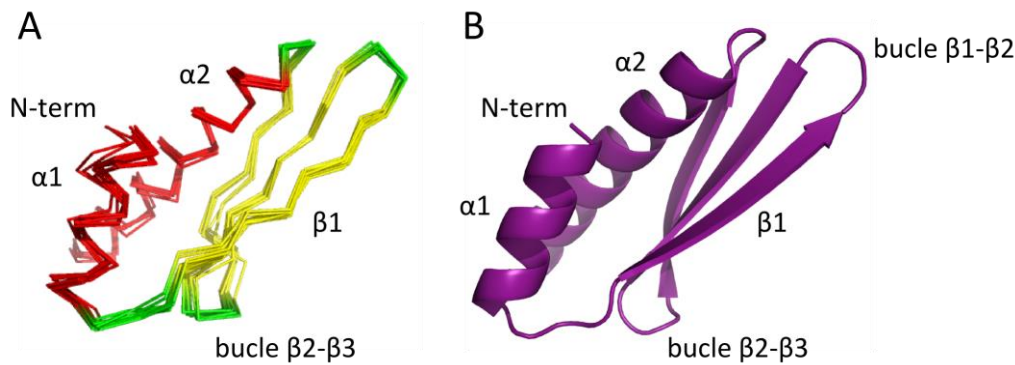
#### 5.1.3. Cálculo de la estructura de DCL1-A en complejo con ARNdh

El cálculo de estructura terciaria de proteínas en solución por RMN se basa principalmente en el uso de un gran número de restricciones de distancia provistas por los NOEs entre núcleos cercanos. Recientemente se ha demostrado que para proteínas de tamaño pequeño se puede obtener una estructura en solución utilizando como restricción únicamente los desplazamientos químicos de la cadena principal (Co, Ca, Cb, Ha, N y HN). Estos valores dependen fuertemente de los ángulos diedros del esqueleto carbonado. Acoplando el método Rosetta para el cálculo *ab initio* de estructuras de proteínas a partir de secuencia con el programa SPARTA para el cálculo de desplazamientos químicos a partir de estructuras se desarrolló el método CS-Rosetta (Shen et al. 2008). Con este método se ha logrado calcular estructuras en solución, a partir de estas restricciones mínimas, que muestran RMSDs menores a 1Å respecto de las estructuras calculadas utilizando un conjunto completo de datos de NOE.

El primer paso para el cálculo de una estructura empleando Rosetta es la selección de fragmentos polipeptídicos a partir de una base de datos de proteínas de estructura conocida. En el protocolo estándar los fragmentos son seleccionados en base a la similitud de su secuencia con la de la proteína en estudio. CS-Rosetta incorpora a esta selección un criterio de similitud entre valores de desplazamientos químicos de la base de datos y valores experimentales. La incorporación de esta información al cálculo mejora notablemente la exactitud de la estructura obtenida (Shen et al. 2008).

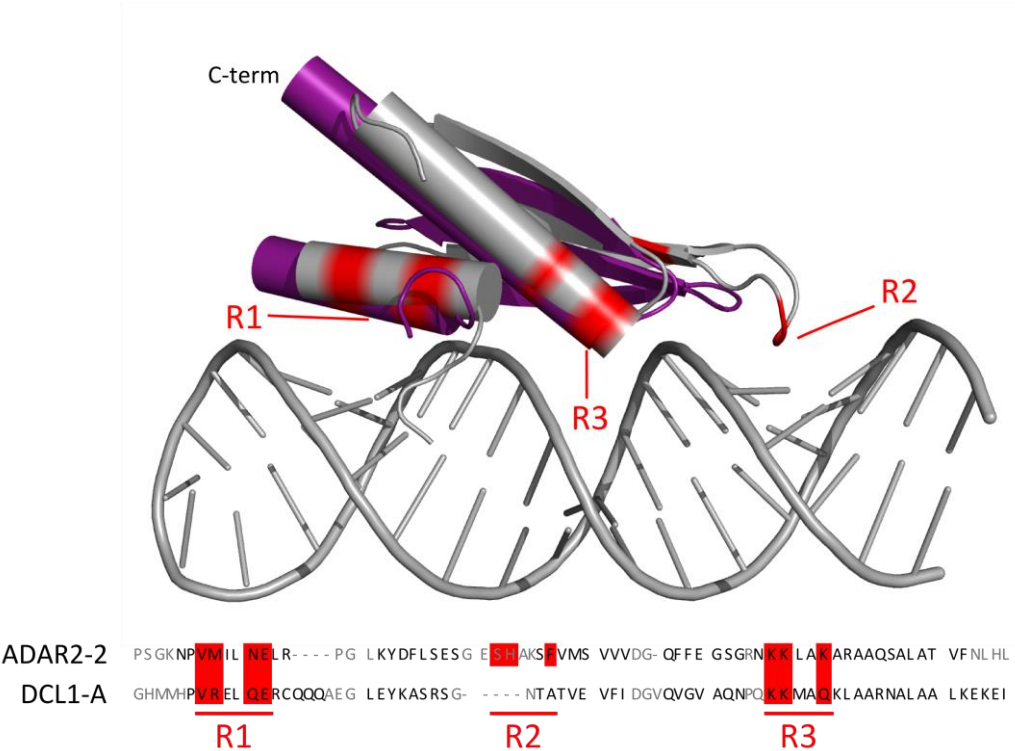
Con los datos de los desplazamientos químicos de los núcleos de la cadena principal de la forma plegada de DCL1-A (apartado 5.1.1) procedimos a realizar el cálculo de la estructura con CS-Rosetta. Sobre 3.500 estructuras calculadas, las 10 de menor energía convergen con un RMSD de  $1,14 \pm 0,38$  Å (Figura 5.3, A). Los elementos de estructura secundaria se condicen con la estructura secundaria calculada a partir de los desplazamientos químicos con SSP en el apartado 5.1.2 (Figura 5.2). La disposición en el espacio de los elementos de estructura secundaria se

corresponde efectivamente con la topología típica de los dsRBDs, las dos hélices alfa sobre las tres hebras extendidas formando una lámina beta antiparalela (Figura 5.3, B). El extremo N-terminal de la hélice 1 presenta mayor variabilidad. Esto se debe a la ausencia de datos experimentales de los desplazamientos químicos para esta región en el cálculo. Recíprocamente, la falta de datos en esta región se debe a que no fue posible medirlos por ausencia de las señales en los espectros (Figura 5.2). La ausencia de las señales se debe probablemente a que los residuos se encuentran en intercambio intermedio entre conformaciones heterogéneas. Por lo tanto, podemos concluir que esta particularidad del modelo se condice con las observaciones experimentales.



**Figura 5.3 DCL1-A adquiere un plegamiento tipo dsRBD en presencia del sustrato.**

A. Alineamiento de las 10 estructuras de menor energía obtenidas para DCL1-A en complejo con el sustrato calculadas con CS-Rosetta. Rojo: hélice alfa, Amarillo: lámina beta, Verde: bucle. B. Modelo de la estructura de menor energía denotando los elementos de estructura secundaria (B).



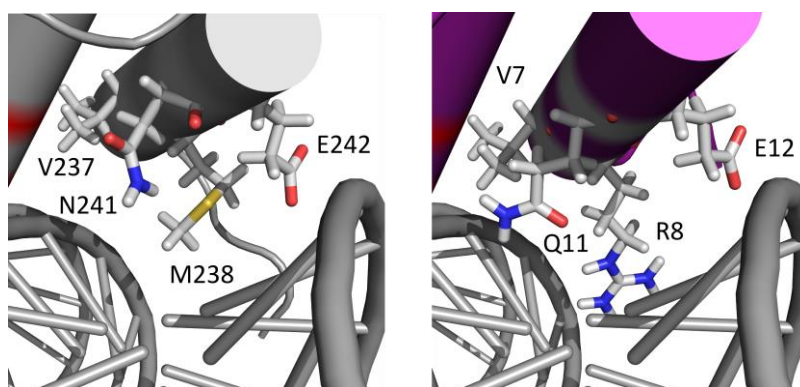
**Figura 5.4 Alineamiento estructural de DCL1-A con el dsRBD-2 de ADAR2.**

En gris se muestra la estructura obtenida en solución por RMN del dsRBD-2 de ADAR2 de rata en complejo



con un fragmento de ARNdh correspondiente a una región del sustrato natural (2L2K.pdb) (Stefl et al. 2010). En púrpura la estructura calculada para DCL1-A alineada con ADAR2-2. En rojo se destacan las regiones de ADAR2-2 involucradas de la interacción con el ARN.

Entre las estructuras conocidas de dsRBDs, sólo algunas han sido determinadas en complejo con el ARN. Por cuestiones experimentales en la mayoría de estas estructuras el ARN empleado consiste en fragmentos cortos de ARN doble hebra apilados en forma axial. Por este motivo, en dichas estructuras los dsRBDs interaccionan con regiones de ARN doble hebra que provienen de dos dúplex de ARN distintos. Consideramos de mayor interés las estructuras de complejos con moléculas únicas tipo hebilla de ARN, por ser estos sustratos más cercanos a los naturales. Por alineamiento estructural entre este grupo de estructuras y nuestra proteína de interés encontramos que el complejo dsRBD:ARN que más se asemeja a DCL1-A es el dsRBD-2 de ADAR2 de rata en complejo con un fragmento de su sustrato (2L2K.pdb) (Stefl et al. 2010). En la Figura 5.4 mostramos el alineamiento estructural de DCL1-A con esta estructura. Observamos que el posicionamiento de las regiones 1 y 3 de unión al ARN (rojo), que se encuentran en las hélices, coincide en ambas estructuras. Por el contrario, la región 2, ubicada en el bucle  $\beta 1$ - $\beta 2$ , es esencialmente diferente.



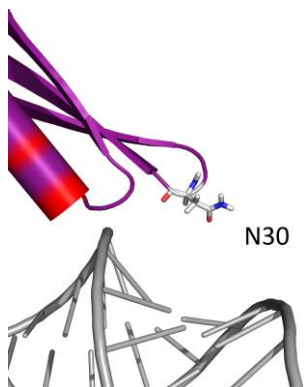
**Figura 5.5** Región 1 de interacción con el ARNdh.

*Izquierda: detalle de la región 1 de interacción con el sustrato de ADAR2-2 (2L2K.pdb), los residuos involucrados se muestran en líneas. Derecha: detalle del alineamiento estructural de DCL1-A con ADAR2-2 en complejo con el sustrato, eliminando ADAR2-2 de la imagen. Los residuos de DCL1-A que coinciden con los de interacción de ADAR2-2 se muestran en líneas. Se destaca la orientación de la R8 de DCL1-A, que podría realizar interacciones tipo puente de hidrogeno con las bases nitrogenadas. Código de color: azul: Nitrógeno; rojo: Oxígeno; amarillo: Azufre; gris: Carbono e Hidrogeno.*

La región de interacción 1 se encuentra en la hélice alfa 1 y reconoce el surco menor del ARNdh. La conservación de la identidad de los residuos involucrados en la unión al sustrato en esta región entre los distintos dsRBDs de estructura conocida es menor en comparación con el resto de las regiones. Esta característica ha llevado a concluir que las propiedades de unión a ARNdh de los distintos dsRBDs está modulada por el carácter fisicoquímico que presentan estos residuos (Masliah et al. 2013). Los residuos que interaccionan con el ARN en ADAR2-2 son V237, M238, N241 y E242, señalados en rojo en la Figura 5.4 y en detalle en la Figura 5.5, izquierda. V237 interacciona por fuerzas de Van der Waals con un residuo de ribosa, mientras que N241 y E242 forman puentes de Hidrogeno con los 2'OH de otros residuos de ribosa a ambos lados del surco menor. A M238 se le ha atribuido la particularidad de establecer un contacto hidrofóbico secuencia específico con la base nitrogenada de un residuo de Adenina (Stefl et al. 2010).

Como podemos ver en el alineamiento de la Figura 5.4, la región 1 de ADAR2-2 coincide con una región homóloga de DCL1-A. Esto nos lleva a suponer que **probablemente los residuos involucrados en la interacción de DCL1-A con el sustrato en esta región sean V7, R8, Q11 y E12** (Figura 5.5, derecha). V7 y E12 se encuentran en la misma posición en la secuencia y muestran una orientación muy similar a los equivalentes en ADAR2-2. N241 de ADAR es reemplazada por Q11. Este cambio no sólo es relativamente conservativo, sino que una glutamina en esta posición es muy frecuente en los dsRBDs en general. Incluso se espera que al tener una cadena lateral más larga, la glutamina interaccione mejor con el sustrato (Masliah et al. 2013). El residuo que nos resulta de mayor interés es R8 de DCL1-A, que se encuentra ocupando el lugar de M238 de ADAR. Como detallamos antes, este es el residuo que realiza el reconocimiento secuencia específico en ADAR. En el caso de DCL1-A podemos ver que **la orientación de R8 sugiere que podría estar realizando contactos con las bases nitrogenadas que permitirían un reconocimiento de la secuencia.**

La segunda región de interacción con el sustrato se encuentra en el bucle  $\beta 1\text{-}\beta 2$  y el contacto se establece con el surco menor una vuelta de hélice más adelante respecto de la región 1. En la Figura 5.4 observamos que este bucle en DCL1-A es más corto que el de ADAR2-2, lo cual elimina la región de unión al sustrato canónica. El bucle  $\beta 1\text{-}\beta 2$  corto es una característica que se repite en los homólogos más cercanos de estructura conocida mencionados en el apartado 3.2.1, los dsRBDs de Dicer de ratón y *K. polysporus*. En el caso de *K. polysporus* se ha sugerido que esta particularidad podría explicar la baja afinidad de este dominio por el ARNdh (Weinberg et al. 2011). Este no sería el caso para DCL1-A ya que la afinidad del dominio por el ARNdh está en el orden esperado para un dsRBD canónico (sección 3.3). Observando en detalle el bucle en cuestión, encontramos que **N30 podría ser un candidato a establecer contactos con el sustrato** (Figura 5.6).

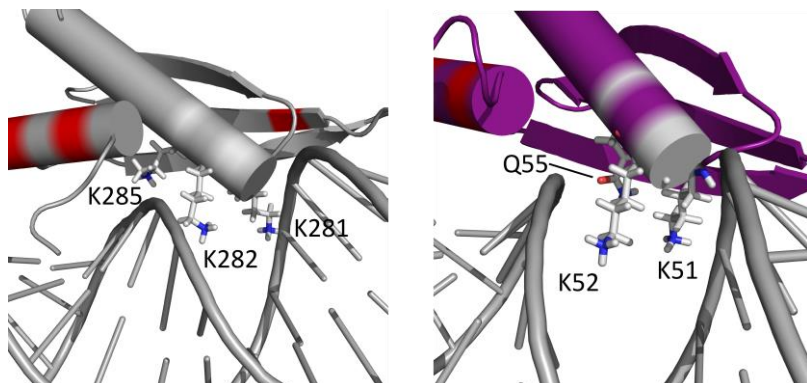


**Figura 5.6 Región 2 de interacción con el ARNdh.**

*A pesar de que el bucle  $\beta 1\text{-}\beta 2$  de DCL1-A es más corto respecto al de ADAR2-2, la orientación de la N30 en principio podría permitir la interacción con el sustrato. Código de color: azul: Nitrógeno; rojo: Oxígeno; gris: Carbono e Hidrógeno.*

**La región 3 se encuentra bien conservada en DCL1-A.** Esta región está ubicada en la hélice 2 e interacciona con el surco mayor de la hélice. El motivo conservado consiste en la secuencia KKxAK, y es uno de los determinantes más relevantes en la diferenciación entre la hélice A del ARNdh y la B del ADN. La segunda y tercer lisina forman puentes de hidrógeno con átomos de oxígeno de los grupos fosfato del esqueleto de una de las hebras de ARN, mientras que la primer lisina lo hace con la hebra complementaria. Este “puente” mide el

ancho del surco mayor de la hélice. En el caso de ADAR2-2 los residuos involucrados en esta región son K281, K282 y K285 (Figura 5.7, izquierda). Los residuos equivalentes en DCL1-A son K51, K52 y Q55.



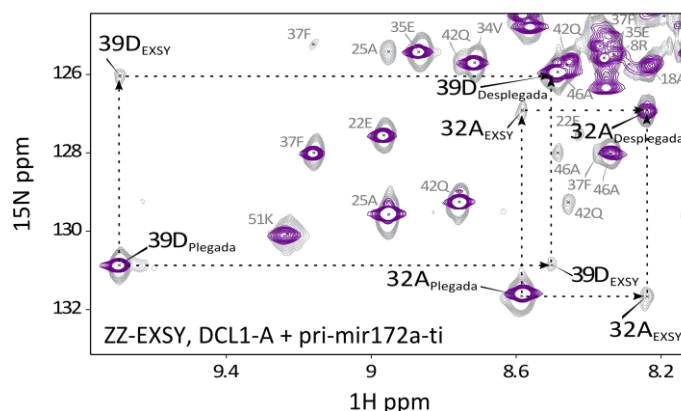
**Figura 5.7** Región 3 de interacción con el ARNdh.

*Izquierda: detalle de la región 3 de interacción con el sustrato de ADAR2-2 (2L2K.pdb), los residuos involucrados se muestran en líneas. Derecha: detalle de la misma región en el alineamiento estructural de DCL1-A con ADAR2-2 en complejo con el sustrato, eliminando ADAR2-2 de la imagen. Los residuos de DCL1-A que coinciden con los de interacción de ADAR2-2 se muestran en líneas. Código de color: azul: Nitrógeno; rojo: Oxígeno; gris: Carbono e Hidrógeno.*

## 5.2. Mecanismo de unión al sustrato y plegamiento

### 5.2.1. DCL1-A plegada se encuentra en intercambio con la forma desplegada

Durante el análisis de los espectros de asignación del complejo observamos que las señales de la proteína desplegada no desaparecen en presencia de exceso de ARNdh. Para confirmar que no estábamos en presencia de una fracción inactiva de la proteína adquirimos un espectro ZZ-EXSY, en el cual se observan picos de cruce entre señales que están en intercambio conformacional entre sí. Realizamos este experimento sobre la misma muestra que utilizamos para la asignación del esqueleto peptídico (apartado 5.1.1).



**Figura 5.8** Espectro ZZ-EXSY del dominio DCL1-A en presencia de pri-miR172a-ti.

*Detalle de la superposición de un espectro HN-TROSY (púrpura) con un espectro ZZ-EXSY (gris) adquirido con un tiempo de mezcla de 200 ms. Se destacan como ejemplo los casos de los residuos 39 y 32, para los cuales se señalan los picos de cruce correspondientes a las resonancias del par HN del grupo amida del enlace peptídico en la forma plegada y desplegada de la proteína, y los picos de cruce etiquetados como*

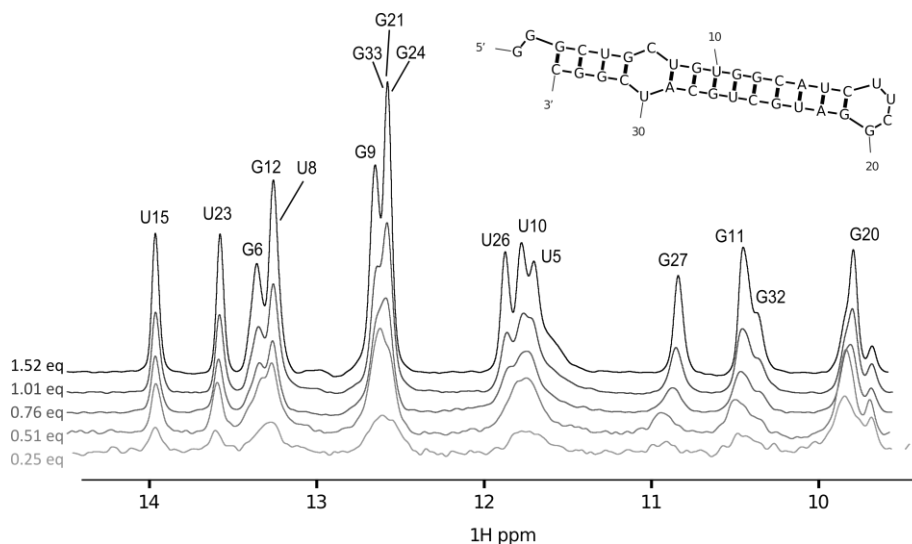
“EXSY” que indican el intercambio entre ambas especies.

Como se observa en la Figura 5.8, pudimos confirmar que la forma desplegada se encuentra en intercambio conformacional con la forma plegada ya que se obtienen los patrones rectangulares de picos de cruce esperados, que se destacan en la figura. De este experimento podemos decir que **la forma desplegada presente en la mezcla es funcional**. El tiempo de intercambio entre la conformación desplegada y plegada está en el rango de 100ms - 1s.

### 5.2.2. La forma desplegada de DCL1-A no desaparece en presencia de exceso de sustrato

Con la intención de obtener información más detallada sobre el mecanismo de unión al sustrato y plegamiento de DCL1-A, realizamos una titulación de la proteína con ARNd<sub>h</sub>. A una muestra 200  $\mu$ M de DCL1-A marcada con  $^{15}$ N agregamos alícuotas de una solución stock 400  $\mu$ M de pri-miR172a-ti. Luego de cada agregado de sustrato adquirimos un espectro HN-HMQC y un espectro unidimensional de protones.

En los espectros de protones es posible seguir la evolución de las señales correspondientes a los protones imino de los pares de base del ARNd<sub>h</sub> (Figura 5.9). La identidad de las señales fue asignada previamente empleando la técnica de la caminata NOESY imino estándar para asignación de ácidos nucleicos. Podemos observar que no ocurren grandes cambios en los desplazamientos químicos de las señales, lo cual nos indica que no ocurren modificaciones significativas en la estructura de la hebilla de ARN. También destacamos que a partir de 1 equivalente, las señales del ARN se vuelven angostas, indicando la presencia de ARN libre en la mezcla. Podemos tomar esto último como un indicio de que **el complejo formado entre DCL1-A y el fragmento pri-miR172a-ti presenta una estequiometría 1:1**.

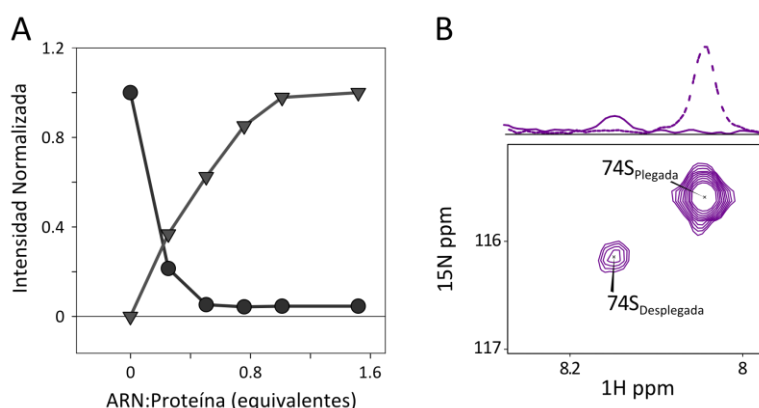


**Figura 5.9** Región de protones imino del fragmento pri-miR172a-ti.

Cada espectro corresponde a agregados consecutivos de pri-miR172a-ti sobre la muestra de DCL1-A, los equivalentes ARN:Proteína de detallan a la izquierda. Destacamos que no ocurren grandes cambios en el desplazamiento químico de las señales.

En los espectros HN-HMQC seguimos la evolución de la proteína. A lo largo de la titulación observamos el aumento progresivo de la intensidad correspondiente a las señales de la forma plegada y la disminución de la de las señales de la forma desplegada (Figura 5.10, A). Este

comportamiento es el esperado para sistemas en intercambio lento, lo cual implica que ambas especies dan origen a un conjunto de señales características de su estado conformacional y el intercambio entre las dos conformaciones es lento en relación a la escala de tiempo de la duración del experimento de RMN. En este régimen de intercambio la intensidad de la señal está relacionada con la cantidad de la especie presente en la mezcla. En este caso podemos observar que una vez saturado el equilibrio a 1.5 equivalentes ARN:Proteína la relación de concentración entra la forma plegada y la desplegada es de 5 a 1 (Figura 5.10, B). De esta manera volvemos a observar el fenómeno descrito en el apartado anterior (5.2.1). La forma desplegada se encuentra presente aún luego del agregado de exceso de sustrato.

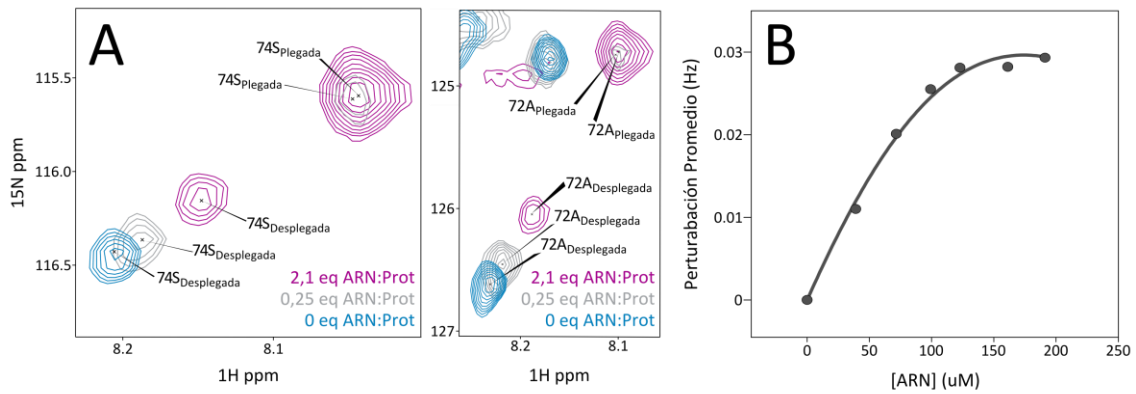


**Figura 5.10** Titulación de DCL1-A con pri-miR172a-ti.

A. Evolución de la intensidad promedio normalizada de las señales correspondientes a todos los residuos asignados de la especie plegada (triángulos) y desplegada (círculos) respecto del agregado de pri-miR172a-ti. B. Detalle del espectro HN-HMQC del punto a 1.5 eq ARN:Proteína, donde se denota la relación 5 a 1 entre la señal de la forma plegada (traza superior, línea interrumpida) y la desplegada (traza superior, línea continua).

### 5.2.3. El mecanismo de unión al sustrato y plegamiento involucra una especie intermediaria

Comparando los espectros HN-HMQC adquiridos a concentraciones crecientes de ARNdh observamos el comportamiento de las señales de la forma desplegada. Como se muestra en la Figura 5.11, A, las señales de esta especie no sólo disminuyen en intensidad, sino que se desplazan hacia una nueva posición en el espectro. Este comportamiento indica que **la conformación de la forma desplegada sufre una alteración en presencia del sustrato**. El desplazamiento progresivo de las señales indica un régimen de intercambio rápido, es decir, distintas especies conformacionales se encuentran en intercambio rápido entre sí, generando una señal única cuyo desplazamiento químico es el promedio entre los desplazamientos químicos originados a partir de cada una de las especies contribuyentes. En este tipo de intercambio la magnitud de la perturbación de los desplazamientos químicos se relaciona con la proporción de cada especie presente en la mezcla. Dado que esta perturbación de los desplazamientos químicos de las señales de la forma desplegada se produce con el agregado de ARNdh (Figura 5.11, B), podemos inferir que **estamos ante la presencia de una tercera especie, en la cual la proteína se encuentra unida al sustrato formando un intermediario previo al evento de plegamiento**. Dado el régimen de intercambio rápido se espera que la afinidad de la forma desplegada por el sustrato sea baja para este paso del mecanismo.

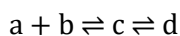


**Figura 5.11** Perturbación de la forma desplegada de DCL1-A con el agregado de pri-miR172a-ti.

A. Detalle de la superposición de espectros HN-HMQC de DCL1-A en presencia de 0, 0,25, y 2,1 equivalentes de ARNdh sustrato. Se indican las señales correspondientes al par HN amida de los residuos 74 y 72, destacando que la señal correspondiente a la forma plegada aumenta en intensidad y se mantiene en la misma posición, mientras que la señal de la forma desplegada disminuye en intensidad y se desplaza. B. Perturbación promedio de los desplazamientos químicos de todos los residuos asignados a la forma desplegada en función del agregado de ARNdh, datos experimentales (círculos) junto a curva teórica en base a ecuaciones de equilibrio (línea) estimando una  $K_D = 2 \mu M$  para el paso de unión al sustrato.

Integrando la información obtenida sobre el mecanismo de formación del complejo podemos proponer el modelo ilustrado en la Figura 5.12, en el cual **el primer paso es la formación de un complejo intermediario con baja afinidad de la forma desplegada con el sustrato, que luego sufre un reordenamiento conformacional que da como resultado el complejo de la forma plegada unida a ARNdh.**

Para completar la descripción del mecanismo propuesto planteamos uno modelo que permite cuantificar las constantes de equilibrio involucradas:



Donde:

a = ARN libre

b = Proteína libre desplegada

c = Complejo intermediario de proteína desplegada unida al ARNdh

d = Complejo final de proteína plegada unida al ARNdh

Constantes de equilibrio:

$K = c/ab$ ; constante de equilibrio del primer paso

$J = d/c$ ; constante de equilibrio del segundo paso

Concentraciones:

$$A = a + c + d; \text{ARN total}$$

$$B = b + c + d; \text{Proteína total}$$

$$a = A - c - d$$

$$b = B - c - d$$

$$ab - c/K = 0$$

$$(A - c - d)(B - c - d) - c/K = 0$$

En las condiciones de los experimentos de RMN:  $A \approx B \gg 1/K$ .

Las concentraciones de todos los componentes pueden ser calculadas de los valores de A, B, K y J con ecuaciones cuadráticas.

Resolviendo para c:

$$d = Jc$$

$$(A - c - Jc)(B - c - Jc) - c/K = 0$$

$$\frac{AB}{(1+J)^2} - c \left[ \frac{(A+B)}{(1+J)} - \frac{1}{K(1+J)^2} \right] + c^2 = 0$$

La evolución de c corresponde a la perturbación de los desplazamientos químicos de las señales en intercambio rápido de la forma desplegada durante la titulación seguida por RMN. De esta titulación conocemos también los valores de A y B. Fijamos el valor de J en 5, por la relación de las intensidades de las señales entre la forma plegada y desplegada en el punto final de la titulación (Figura 5.10, B). De esta manera, podemos obtener una estimación del valor de K del ajuste no lineal a la ecuación cuadrática para la evolución de c de los datos experimentales de la perturbación de los desplazamientos químicos de la forma desplegada en la Figura 5.11, B. Este ajuste arroja un valor de  $K = 0.5 \mu\text{M}$ . Podemos observar la buena correlación del ajuste comparando la curva en el gráfico del panel B de la Figura 5.11 con los puntos experimentales.

Por otro lado, los ensayos de anisotropía de la fluorescencia se llevan a cabo en condiciones en las cuales  $B < K \ll A$ . En estas condiciones  $a \approx A$ . Lo que se mide en los ensayos de anisotropía de fluorescencia es la relación entre ARN unido (anisotropía alta) y ARN libre (anisotropía baja), que es  $(c + d)/B$ . Considerando las constantes de disociación  $K_D = 1/K = ab/c$  y  $J_D = 1/J = c/d$ , obtenemos:

$$B = b + \frac{Ab}{K_D} + \frac{Ab}{K_D J_D}$$

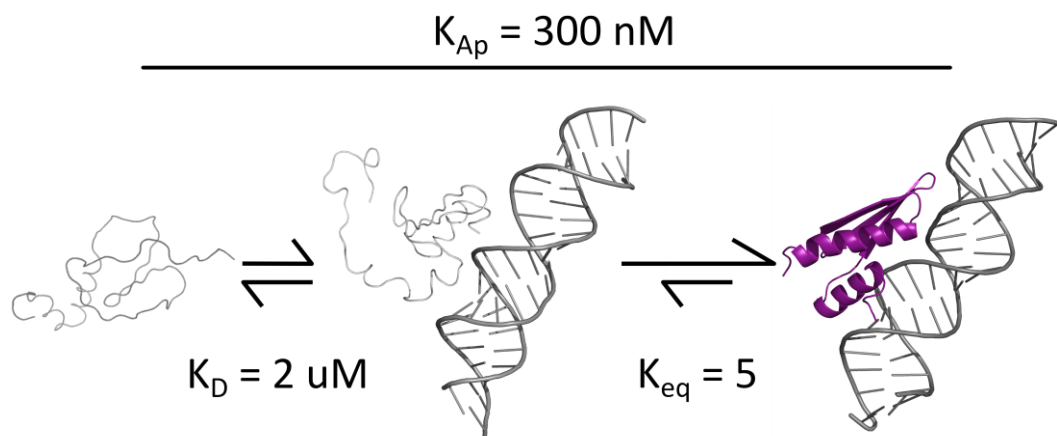
Desde la cual llegamos a:

$$\frac{c + d}{B} = \frac{A}{\frac{K_D J_D}{J_D + 1} + A}$$



Esto significa que en los ensayos de anisotropía de la fluorescencia obtenemos una constante de disociación aparente:  $K_{ap} = K_D J_D / (J_D + 1)$ . De los valores obtenidos de los experimentos de RMN sabemos que  $K_D = 1/K = 1/0.5 \mu M = 2 \mu M$  y  $J_D = 1/J = 1/5 = 0.2$ . Por lo tanto, el valor de  $K_{ap}$  es 330 nM. Este valor de condice muy bien con el obtenido experimentalmente ( $300 \pm 50$  nM), apoyando la exactitud de nuestro modelo.

A modo de resumen, incluimos los valores de las constantes de equilibrio estimados en la Figura 5.12.



**Figura 5.12 Modelo de mecanismo de unión y plegamiento de DCL1-A al sustrato.**

La constante macroscópica de 300 nM ( $K_{Ap}$ ) fue obtenida de titulaciones seguidas por anisotropía de la fluorescencia, la constante de  $2 \mu M$  para el primer paso de unión ( $K_D$ ) se obtuvo del ajuste de los datos experimentales de la perturbación del desplazamiento químico de las señales de la forma desplegada con el agregado de sustrato, y la constante de equilibrio de 5 para el paso de plegamiento de la proteína sobre la superficie del ARNdh ( $K_{eq} = J$ ) se obtuvo de la relación entre las intensidades de las formas plegada y desplegada en presencia de exceso de sustrato.

### 5.3. El intermediario del mecanismo de unión explora conformaciones intermedias

Para observar la tendencia a adoptar conformaciones plegadas de las distintas especies involucradas en el mecanismo recurrimos nuevamente a los desplazamientos químicos secundarios y al cálculo de la proporción de estructura secundaria con SSP. Los valores obtenidos de SSP para las tres especies de DCL1-A: desplegada libre, desplegada unida y plegada, se muestran en la Figura 5.13 (gráfico inferior, gris, rojo, y negro respectivamente). Observamos una marcada diferencia entre la forma plegada y las formas desplegadas en la región N-terminal. La forma desplegada unida al sustrato correspondiente al intermediario del mecanismo de unión y plegamiento no adquiere una estructura definida en esta región. Por el contrario, en la región C-terminal observamos que los valores obtenidos para la región correspondiente a la hélice 2 para esta misma especie son intermedios entre la forma libre y la plegada. Basados en estos resultados podemos decir que **al entrar en contacto con el sustrato cambia el conjunto de conformaciones que explora la cadena desestructurada, adquiriendo una tendencia mayor a formar hélice alfa en el extremo C-terminal.**

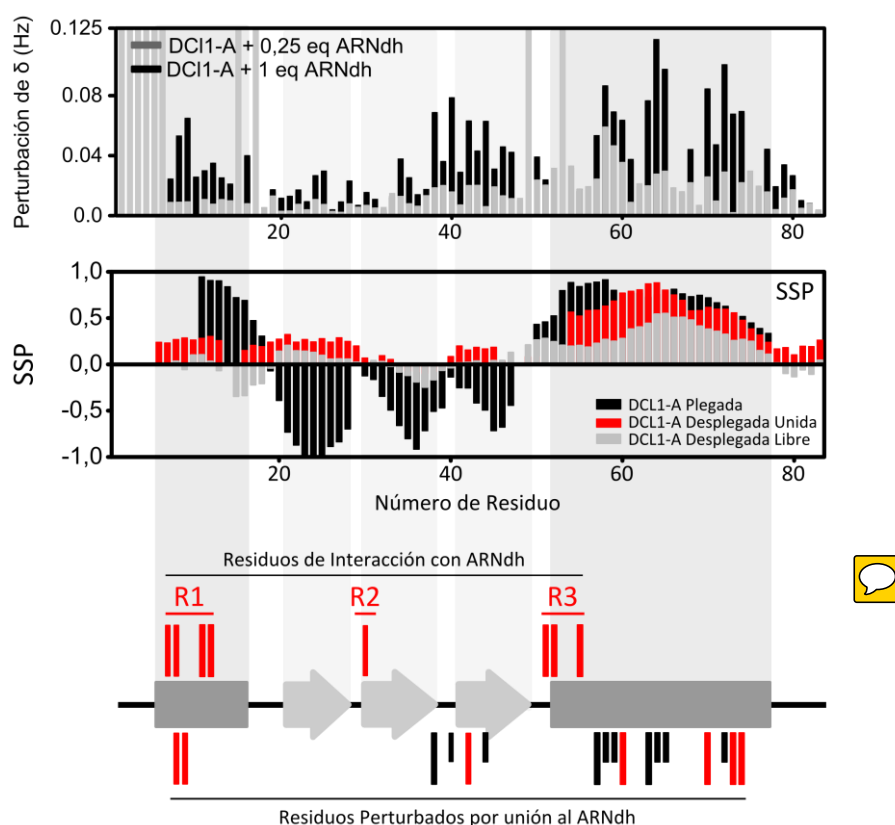
Por otro lado, analizamos la magnitud del cambio en la posición de los picos de los pares HN amida entre las formas desplegada libre y desplegada unida con el agregado de ARNdh (Figura 5.13, Superior). La perturbación de la posición de los picos en el espectro HN-HSQC se relaciona



con un cambio en el entorno del residuo, ya sea por un cambio de conformación como por interacción con otra molécula.

En el extremo C-terminal de la forma desplegada de DCL1-A los residuos más perturbados son **L57, A58, A59, L63, A64, A65, E70, A72, E73 y S74**. Estos residuos se encuentran en la región de la hélice 2, que se estabiliza al formarse el complejo intermediario. Por lo tanto, podemos pensar que **el cambio en el desplazamiento químico podría deberse a la formación de la hélice completa**. De todas maneras no podemos descartar una interacción con el sustrato.

En la región N-terminal observamos que los residuos R8, E9, I38, G40, Q42 y G44 se ven notoriamente perturbados (Figura 5.13, esquema inferior) pero no presentan valores de SSP que indiquen adquisición de estructura secundaria, sugiriendo que su entorno químico/conformacional se ve modificado en este complejo intermediario pero continúan en una conformación desestructurada. El poco cambio en la conformación para estos residuos indica que la perturbación **podría deberse a la interacción con el ARNdh**, especialmente **R8, E9, I38 y Q42** por presentar cadenas laterales largas.



**Figura 5.13** La hélice 2 se estabiliza en el complejo intermediario de DCL1-A con el sustrato.

Gráfico superior: perturbación de los desplazamientos químicos ( $\delta$ ) del par HN amida de cada residuo en presencia de 0.25 (gris) y 1 (negro) equivalentes de ARNdh. Gráfico inferior: proporción de estructura secundaria calculada con SSP (Marsh et al. 2006) de las especies plegada (negro), desplegada unida (rojo) y desplegada libre (gris). Valores positivos de SSP indican conformación de hélice, valores negativos indican conformaciones extendidas o de lámina beta. En el esquema inferior se muestra como referencia los elementos de estructura secundaria de la forma plegada. Las barras rojas superiores corresponden a los residuos putativos de unión al ARNdh. Las barras inferiores corresponden a los residuos perturbados en la formación del complejo intermediario. Las rojas corresponden a residuos polares o cargados, las negras

a residuos apolares. Barras más cortas corresponden a residuos de cadena corta (A y G), mientras que barras más largas corresponden a residuos con cadenas más largas.

Por alineamiento de la estructura de la forma plegada con el dominio ADAR2-2, pudimos inferir cuáles serían los residuos candidatos a estar involucrados en la interacción con el ARNdh en el complejo final (sección 5.1, Figura 5.13, esquema inferior). Comparando estos residuos con los más perturbados durante la formación del complejo intermediario observamos que no coinciden, a excepción de la Arginina 8, como se muestra en el esquema inferior de la Figura 5.13.

Al comparar la naturaleza de los residuos putativos de unión al ARNdh con la de los perturbados en la formación del complejo intermediario (Tabla 5.1), se destaca que los primeros son la mayoría de carácter polar y presentan cadenas laterales extendidas. Entre los perturbados en el complejo intermediario encontramos varias Alaninas y Glicinas que no corresponden con las características fisicoquímicas de los residuos de interacción con el sustrato (Figura 5.13, esquema inferior, barras negras cortas). Observando la distribución de los residuos en el esquema inferior de la Figura 5.13, podemos pensar que el cambio en el entorno de las Alaninas y Glicinas puede ser inducido por la acción de los otros residuos perturbados de cadenas laterales más extendidas. En cuanto a estos últimos podemos pensar que R8, E9, Q42, R60, E70, E73 y S74 podrían ser candidatos a interaccionar con el sustrato por presentar cadenas laterales polares con capacidad de formar puentes de hidrógeno con el ARN. Los residuos con cadenas alifáticas I38, L57 y L63 en principio también podrían establecer contactos hidrofóbicos con las bases nitrogenadas del sustrato. Como describimos previamente, no podemos discernir entre un cambio en la conformación de la cadena y la interacción directa con el sustrato, pero sí podemos decir que los cambios se deben a la presencia del ARNdh y que probablemente algunos de estos residuos estén estableciendo contactos con el sustrato.

En conjunto este análisis nos sugiere que cuando DCL1-A se aproxima al sustrato se forma un complejo en el cual se estabiliza la hélice 2 en la región C-terminal. En este complejo la región N-terminal permanece desestructurada e interaccionando con el sustrato. Los contactos con el ARN podrían estar produciéndose a través de residuos diferentes a los de unión al sustrato esperados. Estas interacciones primarias deberían romperse para dar lugar a la unión específica y el plegamiento de la proteína, explicando la coexistencia de las especies desplegada unida y plegada en el equilibrio y en intercambio lento entre sí.

Tabla 5.1 Residuos involucrados en la unión al sustrato.

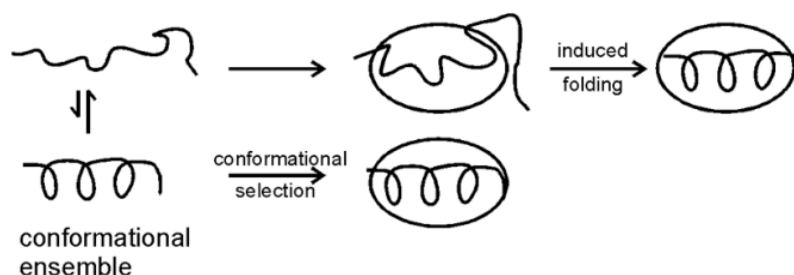
Residuos de unión al ARNdh	Residuos perturbados en la formación del complejo intermediario
V7, R8, Q11, E12, N30, K51, K52, Q55	R8, E9, I38, G40, Q42, G44, L57, A58, A59, R60, L63, A64, A65, E70, A72, E73, S74

Rojo: residuos con cadenas laterales cargadas o polares, negro: residuos con cadenas laterales alifáticas y Glicinas.

## 5.4. Discusión

En este capítulo abordamos la descripción de DCL1-A en complejo con el ARNdh y su mecanismo de formación. La homogeneidad del complejo formado indica que DCL1-A reconoce específicamente al sustrato. Analizando la orientación de las distintas regiones de la proteína plegada logramos identificar que los residuos putativos de interacción con el sustrato se encuentran conservados respecto de otros dominios de la misma familia. Finalmente logramos determinar que el mecanismo a través del cual la proteína reconoce el ARNdh ocurre a través de la formación de un complejo intermediario en el cual el dsRBD se encuentra mayoritariamente desplegado.

Uno de los grandes interrogantes en el campo de las proteínas desestructuradas que se pliegan al interaccionar con el ligando consiste en la elucidación del mecanismo a través del cual se da el plegamiento. Existen dos posibles mecanismos extremos, el plegamiento inducido por la unión o la selección conformacional (Figura 5.14). En el primero, el plegamiento es inducido por los contactos establecidos con el sustrato. En el segundo, la conformación plegada está presente entre las múltiples conformaciones que explora la cadena flexible y es seleccionada por ser esta conformación particular la que permite la unión al ligando (Wright & Dyson 2009).



**Figura 5.14** Las proteínas desestructuradas pueden experimentar dos tipos de mecanismos de unión al ligando y plegamiento.

*El plegamiento puede ser inducido luego de la unión al ligando (mecanismo superior) o la conformación plegada puede ser seleccionada del ensamble conformacional de la forma libre (mecanismo inferior) (Wright & Dyson 2009).*

En el caso de DCL1-A sabemos que en la forma libre se encuentra presente una cierta proporción de estructura secundaria que podría ser esencial para la unión al ligando. Sabemos que el complejo de encuentro con el ligando se produce a través de un estado aún desordenado. Esto nos lleva a la conclusión de que el mecanismo a través del cual se produce la unión y plegamiento no es por selección conformacional de la forma plegada final de la proteína. Este plegamiento final se adquiere una vez unida la proteína al sustrato. Sin embargo, los resultados obtenidos sobre la proteína desplegada indican que la región 3 de interacción con el sustrato se encuentra en un sitio preformado. Esto implica que localmente puede estar operando un mecanismo de selección conformacional. Por lo tanto, podemos describir al mecanismo de unión y plegamiento de DCL1-A como la combinación de los mecanismos posibles. Globalmente estaríamos ante un mecanismo de plegamiento inducido por interacción con el sustrato, mientras que localmente en la hélice 2 se estaría produciendo la selección conformacional de la región 3 de interacción.

Se conocen varios casos en los que una proteína desestructurada forma un complejo de encuentro con su molécula blanco con una velocidad rápida y luego sufre un evento de plegamiento más lento (Wright & Dyson 2009; Narayanan et al. 2008). Uno de los ejemplos lo constituye la unión del dominio desestructurado pKID al dominio KIX. El mecanismo de unión y plegamiento de pKID ocurre a través de la formación de un complejo intermediario en el cual el dominio permanece desestructurado (Sugase et al. 2007; Turjanski et al. 2008). Los contactos entre la forma desplegada y KIX ocurren a través de residuos que finalmente son esenciales para la unión firme del complejo plegado. Por otro lado, se demostró que interacciones inespecíficas en la formación del complejo intermediario aumentan la velocidad directa del evento de unión, por lo tanto se consideran relevantes para el proceso de unión y plegamiento (Turjanski et al. 2008). Este sistema nos permite pensar que en el caso de DCL1-A podemos estar ante un sistema similar, en el que el complejo intermediario está regido por interacciones diferentes a las que finalmente resultan en el plegamiento de la proteína, pero que al mismo tiempo podrían ser esenciales para el evento de encuentro con el sustrato.

## 6. CAPITULO 6: Arquitectura de los precursores de micro-ARNs

Una de nuestras motivaciones principales al comienzo de este trabajo fue decodificar la forma en que el complejo de procesamiento reconoce los sitios de corte sobre los precursores de miARNs. Dado que los precursores en plantas no presentan secuencias ni estructuras globales conservadas, el consenso de la comunidad en la actualidad es que la posición de los cortes se define por determinantes estructurales comunes de sus estructuras de hebilla (Bologna & Voinnet 2014b; Cuperus et al. 2011). Esto nos llevó a estudiar la estructura de los precursores de miARN con la idea de complementar nuestros estudios sobre las proteínas de procesamiento.

Para la liberación del miARN maduro son necesarios al menos dos cortes sucesivos. La actividad tipo Dicer de DCL1 da lugar a una distancia de 21 nucleótidos entre un corte y el siguiente. El registro de los cortes sucesivos está definido por la posición del primer corte. Este registro es el que determina en última instancia la secuencia del miARN liberado, y por consiguiente la especificidad de su ARNm blanco. La definición del registro es lo que hace al primer corte el punto más importante a comprender para elucidar el mecanismo de procesamiento de miARNs.

Como describimos previamente, el procesamiento de los pri-miARNs es mediado por DCL1, asistida por las proteínas de unión a ARNdH HYL1 y SE (Vazquez et al. 2004; Lobbes et al. 2006). Estas tres proteínas co-localizan con los pri-miARN en los cuerpos D en el núcleo celular. La presencia de HYL1 y SE aumenta la eficacia y precisión del corte de los precursores por parte de DCL1 (Kurihara et al. 2006; Song et al. 2007). Los detalles moleculares de la manera en que DCL1 junto a las proteínas accesorias encuentra el sitio del primer corte sobre los precursores largos y altamente variables de plantas no se han establecido aún.

Los precursores en plantas son mucho más heterogéneos que su contraparte en animales. La estructura secundaria predicha para estas moléculas presenta una gran heterogeneidad en cuanto a regiones desapareadas y burbujas. Como describimos en la sección 1.4 de la introducción otro factor que contribuye a la complejidad de este sistema en plantas, respecto de animales, es la existencia de distintos mecanismos de biogénesis de miARNs. Cada precursor es procesado a través de uno entre cuatro mecanismos posibles (Bologna & Voinnet 2014a):

- 1- *Mecanismo corto desde la base hacia el bucle.* Una burbuja interna en la estructura seguida de un tallo inferior de ca. 15 nucleótidos de longitud definen la posición del primer sitio de corte. Este tipo de mecanismo se encuentra en la mayoría de las familias de miARNs en plantas (Mateos et al. 2010; Song et al. 2010a; Werner et al. 2010). A pesar de que el tallo inferior puede contener imperfecciones, la transición desde la burbuja interna al tallo se caracteriza por presentar al menos tres pares de bases que definen el comienzo del tallo inferior del miARN (Bologna et al. 2013). Este mecanismo implica sólo dos cortes, y el segundo se produce a una distancia de 21 nucleótidos respecto del primero.
- 2- *Mecanismo secuencial desde la base hacia el bucle.* El primer corte se define de la misma manera que en el primer caso. Luego son necesarios dos cortes sucesivos para la liberación del miARN maduro, definido entre el segundo y tercer corte (Bologna et al. 2013).
- 3- *Mecanismo corto desde el bucle hacia la base.* En este caso el procesamiento comienza desde el extremo opuesto, el bucle, y requiere sólo dos cortes. La región terminal de

estos precursores presenta una longitud conservada (ca. 42 nucleótidos) y un bucle pequeño (Bologna et al. 2013).

- 4- *Mecanismo secuencial desde el bucle hacia la base*. En este caso son necesarios cuatro cortes sucesivos. Los precursores de las familias de este grupo presentan tallos largos conservados (Bologna et al. 2009; Bologna et al. 2013; Addo-Quaye et al. 2009).

Los estudios que dan origen a estos modelos de procesamiento se basan principalmente en datos de secuenciación masiva de ARN en los que se detectan los intermediarios de procesamiento de los distintos precursores. A partir de los intermediarios de procesamiento es posible definir cuál es el mecanismo al cual es sometido cada precursor. La estructura de los precursores se calcula *in silico* empleando el servidor Mfold (Zuker 2003). Finalmente, agrupando los precursores según su mecanismo de procesamiento, se analizan los posibles determinantes estructurales que presentan en común. De esta manera, a lo largo de los últimos diez años se han intentado encontrar determinantes estructurales sobre los pri-miARNs que definan el mecanismo de procesamiento.

Mfold es uno de los algoritmos más ampliamente empleados para el cálculo de estructuras secundarias de moléculas de ARN a partir de su secuencia (Zuker 2003). Este programa toma la secuencia y la pliega en todas las estructuras secundarias posibles. Para cada conformación suma los parámetros de energía libre asociados a cada elemento de estructura secundaria y le asigna un valor de energía libre total. Luego compara la energía libre de todas las posibilidades y determina cuál es la estructura de menor energía. La base de datos que utiliza el programa consiste en un conjunto de parámetros de energía libre determinados en un trabajo realizado por Turner y colaboradores en 1999 (Mathews et al. 1999). Estos parámetros incluyen valores de energía libre determinados experimentalmente o calculados (según la disponibilidad de datos) para pares de bases canónicas Watson – Crick, pares de base Guanina – Uracilo, estructuras tipo tallo y bucle y pares de bases no canónicos, considerando la dependencia con la secuencia de los alrededores. En la estructura final calculada sólo se representan como pares de bases apareados los pares canónicos Watson – Crick AU y GC, y los pares de bases no canónicos GU, dejando de lado los pares de bases no canónicos restantes.

En el análisis masivo de las estructuras de los precursores en el trabajo realizado por el grupo del Dr. Palatnik (Bologna et al. 2013) se asignó un valor de 1 para las posiciones representadas por Mfold como nucleótidos desapareados y 0 para las posiciones representadas como apareadas AU, GC y GU. Luego, estos valores fueron utilizados para definir la frecuencia de apareamiento de cada posición sobre el alineamiento de los precursores. Este análisis deja de lado la posibilidad de que en la molécula de ARN los pares de bases no canónicos podrían estar apareados (Davis & Znosko 2007; Lemieux & Major 2002), extendiendo las regiones de doble hélice consideradas por Palatnik y colaboradores en su análisis. Esto podría estar dificultando la identificación de determinantes estructurales. Para validar e identificar los determinantes estructurales en precursores encontramos interesante determinar experimentalmente las estructuras secundarias de los pri-miARN.

La determinación de estructuras de moléculas de ARN del tamaño de los pri-miARN es muy difícil de obtener empleando cristalografía o RMN, debido a su elevada flexibilidad en el primer caso y al elevado peso molecular en el segundo. Por esta razón, la forma que elegimos para estudiar a los precursores es a través de técnicas de baja resolución. Las técnicas de mapeo químico y enzimático nos permiten determinar la estructura secundaria y flexibilidad del esqueleto de la molécula de ARN a nivel de cada nucleótido.

Con la intención de validar las características estructurales que determinan el tipo de procesamiento al cual se somete cada tipo de precursor nos propusimos determinar el mapa de estructura secundaria de cinco precursores. Elegimos los precursores a ensayar en base a la disponibilidad de información previa respecto de su mecanismo de procesamiento y evidencia *in vivo* sobre el efecto de mutaciones. En la Tabla 6.1 resumimos los precursores para los cuales determinamos el mapa de estructura secundaria.

**Tabla 6.1 Precursores ensayados**

Precursor	Mecanismo de procesamiento	Determinantes estructurales
Pri-miR172a	Corto desde la base al bucle	Tallo inferior apareado Bucle terminal desestructurado
Pri-miR171a	Corto desde la base al bucle	Tallo inferior apareado Bucle terminal desestructurado
Pri-miR171b	Corto desde el bucle a la base	Tallo terminal apareado
Pri-miR319a	Secuencial desde el bucle a la base	Tallo terminal apareado
Pri-miR394a	Secuencial desde la base al bucle terminal	Tallo inferior apareado

## 6.1. Mapeo estructural por digestión enzimática

Una de las técnicas disponibles para determinar la estructura secundaria de moléculas de ARN es la que explota la existencia de enzimas ARNasas que digieren selectivamente el ARN de forma estructura específica. La Nucleasa V1 corta el ARN en regiones donde la hebra se encuentra en conformación de hélice. El sitio de reconocimiento del sustrato de la enzima necesita entre 4 y 6 nucleótidos consecutivos en dicha conformación para poder llevar a cabo su acción. La conformación de hélice puede derivar del apareamiento de dos hebras complementarias o del apilamiento de las bases nitrogenadas consecutivas en una hebra simple de ARN (Lowman & Draper 1986; Chaulk et al. 2014). Por otro lado, la Nucleasa S1 digiere selectivamente regiones simple hebra sin importar la conformación de la hebra (Ehresmann et al. 1987). S1 posee actividad tanto exo como endonucleolítica. En combinación estas dos enzimas nos permiten obtener un mapa de la estructura secundaria en la que se encuentra cada región de los precursores de miARN, como resumimos en la Tabla 6.2.

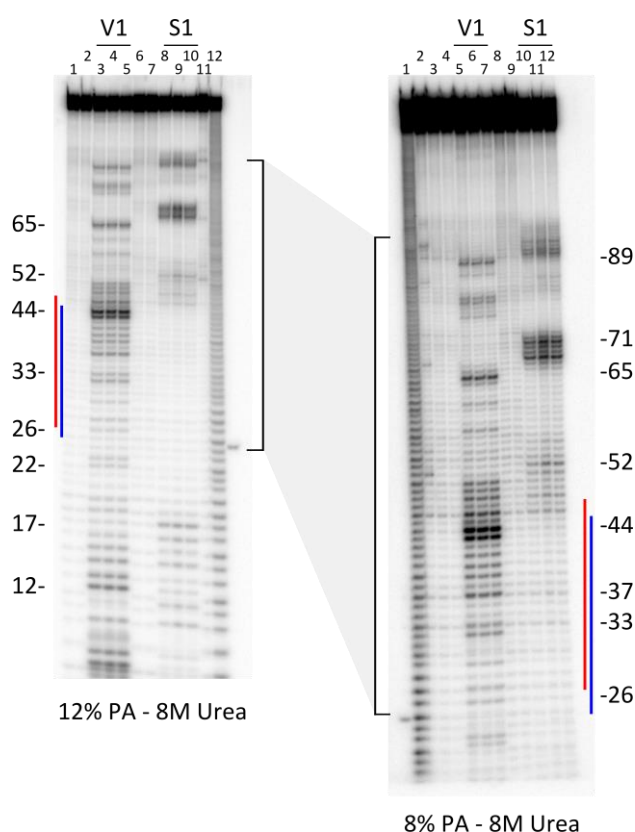
**Tabla 6.2 Análisis de los resultados de digestión enzimática de ARN**

Digestión por Nucleasa V1	Digestión por Nucleasa S1	Estructura Secundaria	Código de color en Figuras
+	-	ARN doble hebra	Rojo
+	+	ARN simple hebra con bases apiladas	Gris
-	+	ARN simple hebra con bases flexibles	Azul

El corte del ARN se produce por hidrólisis a la altura del grupo fosfato del enlace fosfodiéster de una de las hebras. Ambas enzimas dan como producto dos fragmentos, uno con un extremo

3'OH y el otro con un 5'H<sub>2</sub>PO<sub>4</sub>. Los productos de las digestiones se visualizan resolviendo las bandas por electroforesis en geles desnaturalizantes. Como las muestras de ARN tratadas están marcadas en el extremo 5', las bandas en el gel corresponden a los fragmentos del lado proximal del corte (respecto al 5' de cada molécula). La presencia de bandas nos indica que las enzimas han digerido la muestra, lo cual implica que el esqueleto de la hebra se encontraba en la conformación preferida por cada enzima. La intensidad de cada banda reporta la proporción de estructura adquirida o disponibilidad de cada región en particular para ser procesada por el sitio activo de la enzima.

Para realizar nuestros ensayos produjimos las muestras de cada precursor por transcripción *in vitro* y lo marcamos en el extremo 5' con <sup>32</sup>P. Al momento de realizar los ensayos calentamos las muestras a 95°C y enfriamos inmediatamente en hielo para favorecer el plegamiento por hibridación intramolecular. Luego incubamos a temperatura ambiente durante 15 minutos y agregamos una solución tampón conteniendo iones K<sup>+</sup> para estabilizar la estructura secundaria. A continuación procedimos a la digestión con las distintas enzimas a la misma temperatura. Optimizamos las condiciones para obtener una digestión tal que se produzca un único corte por molécula de ARN. Para visualizar los resultados sembramos las muestras en geles de poliácridamida en condiciones desnaturalizantes, que luego revelamos usando pantallas sensibles a la radiación del <sup>32</sup>P. Finalmente visualizamos las bandas empleando un escáner óptico. Para cuantificar la intensidad de las bandas empleamos el programa SAFA (Laederach et al. 2009).



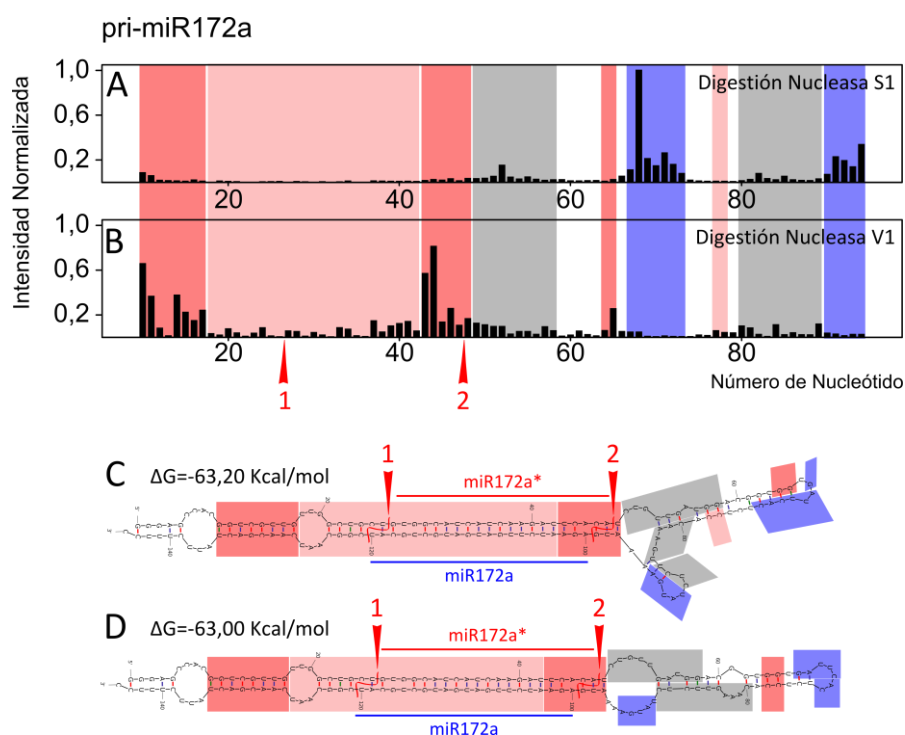
**Figura 6.1 Digestión de pri-miR172a con nucleasas específicas para regiones con distintas estructuras secundarias.**

*Izquierda: gel de poliácridamida 12% urea 8M. Calles 1 y 6: pri-miR172a, calle 2: pri-miR172a + solución tampón de la Nucleasa V1, calles 3, 4 y 5: pri-miR172a + 0,01 U Nucleasa V1, calle 7: pri-miR172a + solución tampón de la Nucleasa S1, calles 8, 9 y 10: pri-miR172a + 1 U Nucleasa S1, calle 11: pri-miR172a*



+ Nucleasa T1, calle 12: pri-miR172a + NaOH. Derecha: gel de poliacrilamida 8% urea 8M. Calle 1: pri-miR172a + NaOH, calle 2: pri-miR172a + Nucleasa T1, calle 3: pri-miR172a, calle 4: pri-miR172a + solución tampón de la Nucleasa V1, calles 5, 6 y 7: pri-miR172a + 0,01 U Nucleasa V1, calle 8: pri-miR172a, calle 9: pri-miR172a + solución tampón de la Nucleasa S1, calles 10, 11 y 12: pri-miR172a + 1 U Nucleasa S1. El contraste de la imagen ha sido modificado para mayor claridad. Las líneas rojas señalan la región correspondiente al miR172a\*. Las líneas azules señalan la región complementaria al miR172a.

En la Figura 6.1 se muestran los geles obtenidos para el precursor pri-miR172a tratado con las Nucleasas V1 y S1. Para poder resolver y cuantificar bandas de un amplio rango de peso molecular corremos dos geles de distinto porcentaje de acrilamida (12 y 8 %). Como referencia usamos una escalera de hidrólisis alcalina, en la que la muestra de ARN es tratada con NaOH para obtener una banda por el corte de la cadena a nivel de cada nucleótido. La segunda referencia es la digestión de la muestra en condiciones desnaturalizantes con la enzima T1, que corta específicamente en guaninas desapareadas. De esta manera obtenemos una banda por cada nucleótido de Guanina en el precursor y el patrón de bandas permite identificar la región de la secuencia a la que pertenecen.



**Figura 6.2** Análisis de los patrones de digestión de las Nucleasas V1 y S1 sobre pri-miR172a.

A. Intensidades normalizadas de las bandas de digestión con Nucleasa S1. B. Intensidades normalizadas de las bandas de digestión con Nucleasa V1. C. Estructura de menor energía obtenida con el servidor Mfold con parámetros establecidos por defecto por los autores (37°C, NaCl 1M) (Zuker 2003). D. Estructura alternativa obtenida con los mismos parámetros. Destacamos que la estructura alternativa es la que corresponde a los patrones de digestión, particularmente en la región hacia el bucle. Gris: región simple hebra con bases apiladas. Rojo: región doble hebra. Azul: región simple hebra. Flechas rojas: sitios donde ocurre el primer y segundo corte por parte de DCL1.

En la Figura 6.2 graficamos la intensidad normalizada de las bandas de los geles de la figura anterior (Figura 6.1). Realizamos la normalización de las intensidades respecto de la banda más intensa y la menos intensa presentes en ambos geles. El análisis de los patrones de digestión define distintas regiones sobre el precursor. En azul marcamos las regiones digeridas

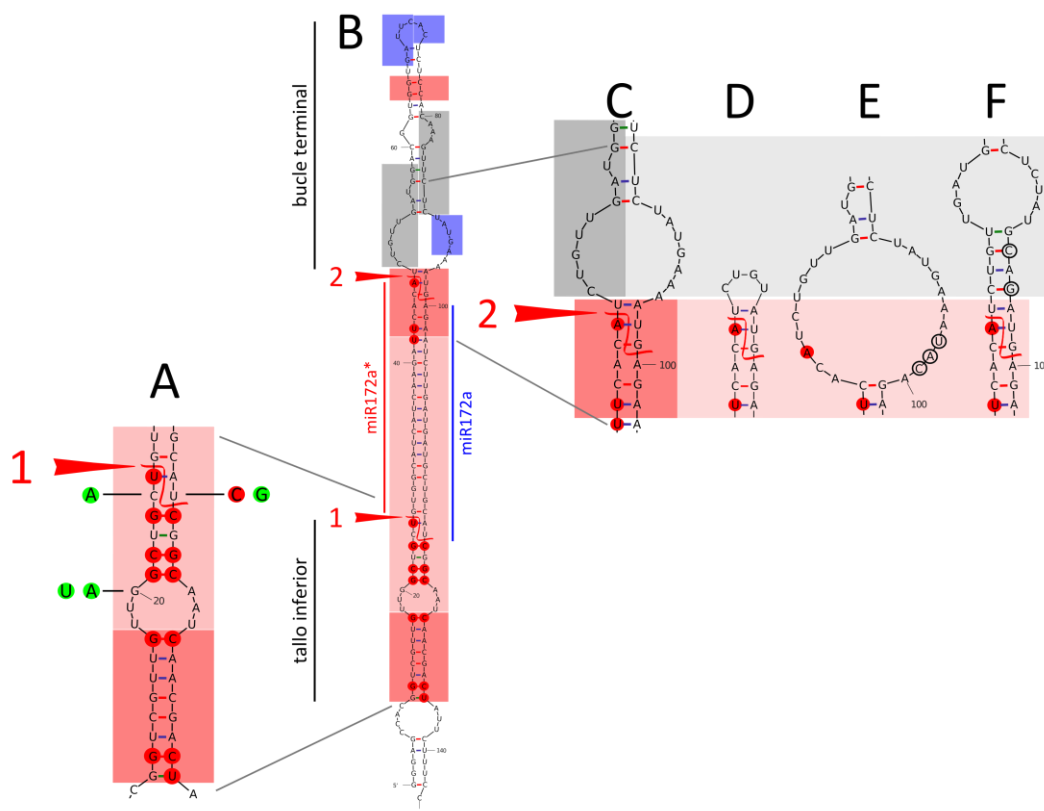
selectivamente por S1 (Figura 6.2, A). En rojo señalamos las regiones donde se produce preferentemente la digestión por V1 (Figura 6.2, B), con mayor y menor intensidad de color según observamos mayor y menor grado de digestión. En gris señalamos una región que es atacada en un grado similar por parte de ambas enzimas. Como describimos anteriormente, esta característica dual de ciertas regiones ha sido descrita anteriormente como regiones donde las bases se encuentran desapareadas pero apiladas por interacciones de van der Waals entre los anillos de las bases nitrogenadas (Chaulk et al. 2014).

Al proceder al análisis por contraste de los patrones de digestión obtenidos con la estructura calculada por minimización de la energía libre según el servidor Mfold (Zuker 2003) encontramos que se presentan algunas diferencias. En la Figura 6.2 C mostramos la estructura de menor energía calculada por Mfold. Esta estructura es la que ha sido típicamente empleada para ilustrar a este precursor y presenta un  $\Delta G$  de -63,2 Kcal/mol. **Un análisis de las estructuras alternativas con energías algo mayores propuestas por el servidor en el mismo cálculo nos lleva a identificar a la estructura mostrada en la Figura 6.2 D como la que responde en mayor medida a los datos obtenidos por la digestión con las enzimas específicas.**

Uno de los determinantes estructurales imprescindible para que se produzca el correcto procesamiento de este precursor es la región denominada tallo inferior, que involucra unos 15 nucleótidos desde el sitio de corte 1 hacia la base (Mateos et al. 2010) (Figura 6.3, B). Por comparación de la estructura calculada para varios precursores que son procesados a través del mismo mecanismo que el pri-miR172a (corto de la base hacia el bucle) se ha establecido que esta región se encuentra preferentemente apareada (Bologna et al. 2013). **A través de nuestros resultados podemos confirmar que la primer mitad del tallo inferior (nucleótidos 10; 11; 14 – 17) presentan un elevado carácter de hélice A, mientras que la segunda mitad presenta un carácter menos marcado (nucleótidos 18 – 26). A pesar de ser relativamente bajo, podemos proponer que el grado de helicidad de esta región es importante para el correcto procesamiento del precursor por contraste de estos resultados con evidencias de procesamiento *in vivo* de mutantes ensayadas por Palatnik y colaboradores (Mateos et al. 2010).** En este estudio se identificaron una serie de mutaciones en esta región, algunas de las cuales impiden el procesamiento y otras son toleradas. En general las toleradas son cambios que conservan la estructura secundaria esperada, en términos de pares de bases tipo Watson – Crick y pares GU (Guanina – Uracilo). Por el contrario, las no toleradas son las que cambian los pares de bases canónicos por pares no canónicos que podrían estar desestabilizando la hélice (círculos rojos). Esto se condice con nuestro mapeo estructural del precursor en el cual confirmamos la estructura tipo hélice del tallo inferior.

En la Figura 6.3 A señalamos un grupo de mutantes que se desvían en cierto grado de esta tendencia. Según el plegamiento predicho por Mfold hay una pequeña burbuja entre los nucleótidos 18 – 20 y sus complementarios 125 – 127. Esta burbuja se produce por el hecho de que Mfold no admite pares de bases no canónicos en su cálculo. En el mapeo por V1 vemos que estas posiciones son atacadas, lo cual nos indica la existencia de un cierto carácter tipo hélice A. Para cerrar la burbuja es preciso formar el par de bases 18U-127U y 20G-125A. Ambos tipos de pares de base suelen encontrarse en estructuras de ARN (Lemieux & Major 2002). Las mutaciones toleradas en la Guanina 20 (círculos verdes) incluyen un reemplazo por un Uracilo, lo cual cerraría el par de bases tipo Watson – Crick con la Adenina 125, estabilizando el carácter de hélice de esta región. Por otro lado, la segunda mutación tolerada introduce un par AA, el cual también se encuentra apareado en otras estructuras de ARN (Lemieux & Major 2002). En

resumen observamos que el carácter tipo hélice A que observamos para esta región se condice con los datos experimentales de procesamiento del precursor *in vivo* (Mateos et al. 2010).



**Figura 6.3 Relación de la estructura con las mutantes conocidas del precursor de miR172a.**

A. Detalle de la región del tallo inferior indicando mutaciones ensayadas *in vivo* en Mateos et al. 2010. Verde: mutaciones toleradas. Rojo: mutaciones no toleradas. B. Mapeo estructural del pri-miR172a obtenido por digestión con nucleasas específicas sobre el cual señalamos las regiones del precursor y en círculos rojos marcamos los sitios relevantes para su correcto procesamiento (adaptación de (Mateos et al. 2010)). C. Detalle de la región terminal sobre el corte 2 de DCL1. D. Mutante carente de bucle terminal. E. Mutante de bucle abierto (3 mutaciones puntuales marcadas con círculo negro). F. Mutante de bucle cerrado (2 mutaciones puntuales marcadas con círculo negro). Tanto D como E y F fueron ensayadas *in vivo* por Mateos et al. 2010. D y E no son procesados por DCL1. F presenta un procesamiento normal.

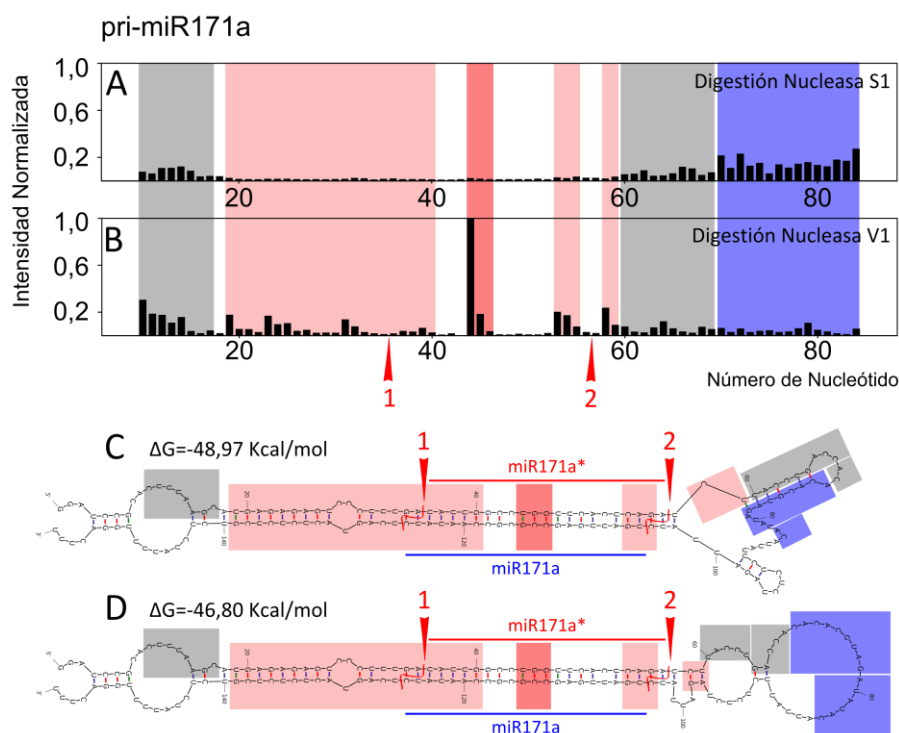
El otro grupo de mutantes que destacamos en la Figura 6.3 A es el que afecta la posición de la Citosina 25, enfrentada al Uracilo 120. Esta posición resulta de gran interés ya que está ubicada en el sitio de corte de DCL1. Según los resultados obtenidos *in vivo* por el grupo del Dr. Javier Palatnik, las mutaciones ilustradas que implicarían un par AU y un CG (círculos verdes), además del reemplazo de ambas bases por un par GG o CA (no ilustrados en la figura, comunicación oral del Lic. Edgardo Bresso) son toleradas en esta posición. La única mutación que impide el procesamiento del precursor es el par CC (círculo rojo). En el gráfico de la intensidad de la digestión con V1 en el panel B de la Figura 6.2 observamos que la digestión en la posición 25 de la secuencia silvestre es prácticamente nula, y se encuentra en el contexto de una región de baja intensidad de digestión. El hecho de que tanto un par AU como CG sean tolerados en este sitio indica que el bajo carácter de hélice de la región es de todas maneras relevante, ya que esperaríamos que estos pares de base Watson – Crick estabilicen esta tendencia. Por otro lado, el par CC se encuentra pobremente representado en estructuras de ARN conocidas (Lemieux & Major 2002). De los pares de bases no canónicos el par CC es el único que puede establecer un

solo puente de hidrógeno entre las bases, siendo el que menos energía aportan a la hibridación entre hebras complementarias (Davis & Znosko 2007). Otra característica de este tipo de par es que al encontrarse dentro de un tallo doble hebra interrumpe la estructura tipo hélice A (Tavares et al. 2009). Todos estos indicios nos llevan a proponer que **la estructura de la región alrededor de la Citosina 25 presenta una estructura “débil” de hélice A ya que no es atacada fuertemente por la nucleasa V1. A pesar de no ser una hélice fuertemente estructurada, el carácter de hélice que observamos es probablemente necesario para que esta región sea aceptada por la maquinaria de procesamiento para realizar el corte 1 sobre el precursor.** Este débil carácter de hélice es probablemente susceptible a la presencia de un par CC, lo cual hace que el precursor no pueda ser procesado. Otra región de interés es la que se encuentra alrededor del sitio donde ocurre el segundo corte por parte de DCL1 (Figura 6.3, C). En el ensayo de mapeo estructural observamos una región de fuerte carácter de hélice A entre los nucleótidos 43 a 48, seguida de una región simple hebra pero con apilamiento de bases. Encontramos de gran interés esta característica ya que coincide con datos experimentales previos. En la Figura 6.3 D, E y F se ilustran mutantes del bucle terminal ensayadas *in vivo* (Mateos et al. 2010). La mutante D carece completamente del bucle terminal y no es procesada por DCL1 correctamente. El miR172a está ausente en las plantas sobreexpresantes de esta construcción, sin embargo se observan intermediarios de procesamiento aberrantes. La mutante E posee tres mutaciones puntuales que desarman los pares de bases canónicos en el sitio del corte 2. Esta construcción tampoco produce miR172a cuando se sobreexpresa *in vivo*. Por último, en la mutante F la secuencia silvestre fue modificada de forma de cerrar la región simple hebra por encima del corte 2. Esta mutante es procesada correctamente y su sobreexpresión en plantas da como resultado la presencia de altos niveles de miR172a. Combinando la información aportada por estas tres mutantes y la estructura determinada por nuestro ensayo de mapeo estructural concluimos que **para que se produzca el correcto procesamiento del pri-miR172 es necesario que la estructura alrededor del sitio de corte 2 se mantenga con su carácter de hélice tipo A.** Un indicio de este tipo fue descrito para el sitio donde ocurre el segundo corte por parte de Dicer sobre el precursor pre-let-7 en *C. elegans*. Se demostró que la proteína inhibitoria del procesamiento de este precursor (LIN 28) lleva a cabo su función induciendo un cambio conformacional en el precursor que abre la hélice en la región del segundo corte (Lightfoot et al. 2011).

Una mirada global del mapa de estructura secundaria del precursor muestra que la región con estructura de hélice A incluye al tallo inferior más la región del miR172a/miR172a\*. Esto se condice con la evidencia *in vivo* y sugiere que la región doble hebra es reconocida y aceptada por el complejo de procesamiento para finalmente llevar a cabo la digestión en los sitios correctos. La longitud del tallo inferior doble hebra evidenciada por la digestión con V1 en el caso de este precursor sería de 17 nucleótidos hasta el extremo 3' del corte.

Repetimos este mismo tipo de ensayo para los precursores pri-miR171a, pri-miR171b, pri-miR319a y pri-miR394a.

En la Figura 6.4 mostramos el mapa estructural obtenido para el precursor del miR171a. De la misma forma que en el caso anterior, la estructura de menor energía obtenida con los parámetros por defecto con el servidor Mfold no representa exactamente el patrón de digestión obtenido. Al mismo cálculo incorporamos dos restricciones de estructura: los nucleótidos 58 y 59 deben estar en una región doble hebra, y los nucleótidos 70 – 76 deben estar en una región simple hebra. Como resultado obtuvimos **la estructura en la Figura 6.4 D, que representa de manera más precisa nuestros datos experimentales.**



**Figura 6.4** Mapeo estructural de pri-miR171a.

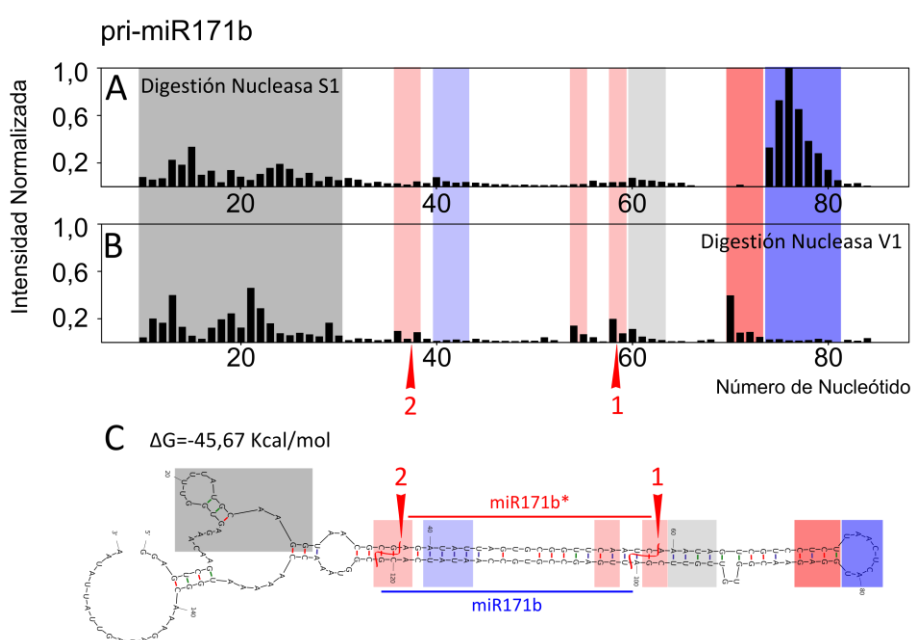
A. Intensidades normalizadas de las bandas de digestión con Nucleasa S1. B. Intensidades normalizadas de las bandas de digestión con Nucleasa V1. C. Estructura de menor energía obtenida con el servidor Mfold con parámetros establecidos por defecto por los autores (37°C, NaCl 1M) (Zuker, 2003). D. Estructura alternativa obtenida incorporando dos restricciones derivadas de las determinaciones experimentales. Incluyendo estas restricciones se logra una mejor representación de los datos experimentales. Gris: región simple hebra con bases apiladas. Rojo: región doble hebra. Azul: región simple hebra. Flechas rojas: sitios donde ocurre el primer y segundo corte por parte de DCL1.

El precursor del miR171a es procesado a través del mismo mecanismo que el del miR172a, corto desde la base hacia el bucle terminal. El determinante estructural es también un tallo inferior apareado (Bologna et al. 2013; Song et al. 2010a). Respecto a esta característica es interesante destacar la zona de digestión mixta entre los nucleótidos 10 – 17 (Figura 6.4, D, gris) que observamos en nuestros ensayos. El patrón de digestión de esta región indica que la estructura es de simple hebra con las bases apiladas. Por lo tanto **la estructura de doble hebra comienza a partir del nucleótido 19, exactamente 17 nucleótidos antes del extremo 3' del sitio del primer corte**. Es decir, la transición simple hebra a doble hebra necesaria para determinar el sitio del corte 1 según nuestros datos experimentales se encuentra a la distancia definida como esencial por Fedoroff y sus colaboradores (Song et al. 2010b), a diferencia de lo que predice la estructura calculada por Mfold.

De la evidencia *in vivo* sobre la importancia de los elementos de estructura secundaria para este precursor sabemos que una mutante en la cual se cerraron las dos burbujas en el tallo inferior inmediatamente por debajo del sitio del corte 1 es tolerada por la maquinaria de procesamiento (Song et al. 2010a). Lo interesante acerca de esta mutante es que la eficiencia del primer corte se ve aumentada, mientras que la del segundo decae respecto del procesamiento del precursor silvestre. Como podemos observar de nuestros ensayos, **el carácter de hélice A del precursor silvestre en esta región es relativamente débil**. El aumento de pares de base tipo Watson – Crick en la mutante mencionada probablemente induzca un

aumento en el carácter de hélice A, alterando posiblemente la interacción con la maquinaria de procesamiento. Los autores del trabajo referenciado consideraron la posibilidad de que al aumentar la proporción de ARN doble hebra aumentara la afinidad de DCL1 por el precursor, haciendo más eficiente la reacción del primer corte pero retardando el rearreglo necesario para llevar a cabo el segundo. Esta propuesta estaría de acuerdo con nuestros datos experimentales.

Con la idea de observar diferencias entre los mapas estructurales de precursores sometidos a diferentes mecanismos de procesamiento procedimos a ensayar al precursor del miR171b. Este precursor se procesa en la dirección opuesta a los dos descritos, desde el bucle terminal hacia la base. El determinante estructural que define el sitio del primer corte es una región doble hebra desde el sitio de corte 1 hacia el bucle terminal (Bologna et al. 2013). El mapa obtenido para este precursor se muestra en la Figura 6.5.



**Figura 6.5 Mapa estructural de pri-miR171b.**

A. Intensidades normalizadas de las bandas de digestión con Nucleasa S1. B. Intensidades normalizadas de las bandas de digestión con Nucleasa V1. C. Estructura de menor energía obtenida con el servidor Mfold con parámetros establecidos por defecto por los autores (37°C, NaCl 1M) (Zuker, 2003). Gris: región simple hebra con bases apiladas. Rojo: región doble hebra. Azul: región simple hebra. Flechas rojas: sitios donde ocurre el primer y segundo corte por parte de DCL1.

Del mapa estructural del pri-miR171b encontramos importante destacar que **la región de marcado carácter doble hebra próxima al bucle terminal define una distancia de exactamente 17 nucleótidos de distancia hasta el extremo 3' del sitio de corte 1**. Esta estructura podría estar determinando el inicio de la región estructurada requerida para la localización del primer corte, de acuerdo con lo establecido por comparación entre estructuras predichas de otros precursores que se procesan mediante el mecanismo desde el bucle a la base (Bologna et al. 2013).

Otro precursor procesado desde el bucle terminal hacia la base es el pri-miR319a, pero en este caso el procesamiento se produce en forma secuencial y son necesarios cuatro cortes para liberar el miARN. El determinante estructural es también una región doble hebra en la región entre el corte 1 y el bucle terminal (Bologna et al. 2009; Bologna et al. 2013). Procedimos a

validar este modelo a través del mismo tipo de ensayo empleado para los otros precursores. En la Figura 6.6 mostramos el mapa obtenido para este precursor.

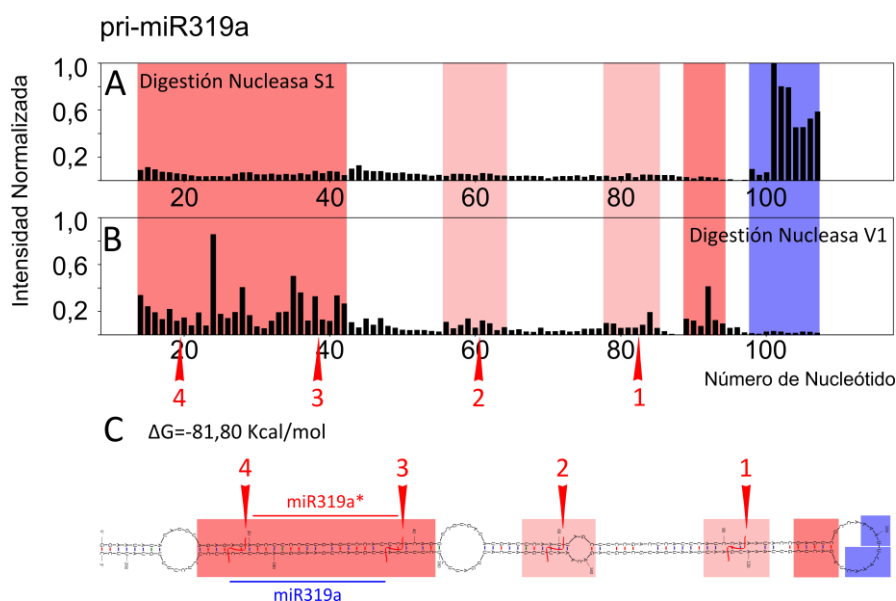


Figura 6.6 Mapa estructural para el pri-miR319a.

A. Intensidades normalizadas de las bandas de digestión con Nucleasa S1. B. Intensidades normalizadas de las bandas de digestión con Nucleasa V1. C. Estructura de menor energía obtenida con el servidor Mfold con parámetros establecidos por defecto por los autores (37°C, NaCl 1M) (Zuker, 2003). Rojo: región doble hebra. Azul: región simple hebra. Flechas rojas: sitios donde ocurren los cortes sucesivos por parte de DCL1.

Nuevamente observamos una región de fuerte carácter doble hebra a una distancia similar a la esperada desde el bucle terminal hacia el sitio de corte 1. En este caso son 14 nucleótidos de distancia desde el comienzo de la región doble hebra hacia el extremo 3' del corte. Mutantes ensayadas *in vivo* en trabajos anteriores indican que esta región es indispensable para el correcto procesamiento del precursor. La eliminación del bucle junto con los últimos 8 pares de bases (nt 86 - 117) impide por completo la acumulación del miARN correcto y resulta en un procesamiento aberrante desde la base hacia el bucle. La eliminación del bucle completo (nt 94 - 109) disminuye la eficiencia de procesamiento pero no altera la precisión de los cortes (Bologna et al. 2009). El conjunto de estos datos indica que **probablemente la región con marcado carácter de hélice A próxima al bucle tenga un papel clave en la definición del sitio de corte 1.**

Otro de los interrogantes respecto de los precursores que sufren un procesamiento secuencial consiste en dilucidar cuál es la característica de dúplex miR/miR\* que hace que la maquinaria de procesamiento continúe realizando cortes hasta llegar a la escisión del miARN correspondiente. Lo que se busca es definir la particularidad de la región que hace que el procesamiento no se detenga anticipadamente o continúe luego de liberado el miARN maduro. Actualmente una propuesta es que la hibridación del dúplex presenta una estabilidad mayor respecto del resto del precursor (comunicación oral del grupo del Dr. Palatnik). A primera vista podemos identificar en nuestro mapa estructural que **la región correspondiente al dúplex miR319a/miR319a\* presenta un carácter de estructura tipo hélice A marcadamente mayor respecto al resto del precursor, lo cual implica una mayor estabilidad del apareamiento entre las**

hebras en esta región. En principio esta particularidad estaría de acuerdo con lo propuesto por el Dr. Palatnik y colaboradores, definiendo la identidad del miARN maduro.

Finalmente realizamos el mismo ensayo para el precursor pri-miR394a, el cual es procesado mediante el mecanismo secuencial pero en dirección desde la base hacia el bucle terminal, al contrario que el pri-miR319a. En la Figura 6.7 mostramos el mapa obtenido. De la misma manera que en casos anteriores la estructura calculada por Mfold no representa exactamente los datos experimentales, por lo tanto procedimos a realizar un segundo cálculo incorporando restricciones en la región del bucle terminal para obtener la estructura de la Figura 6.7 D.

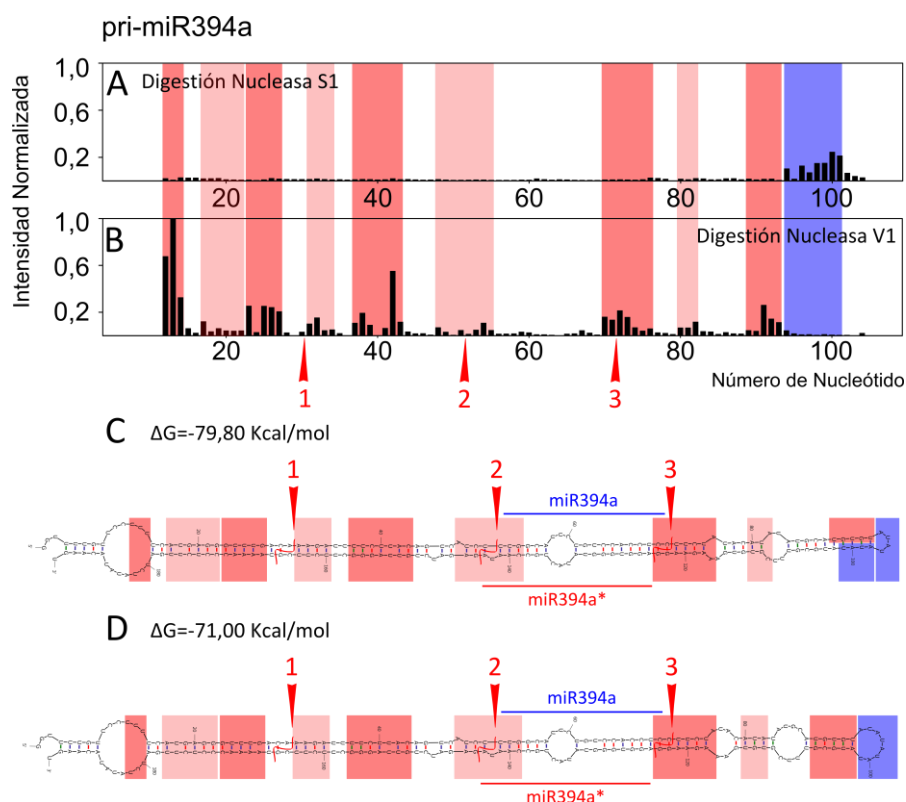


Figura 6.7 Mapeo estructural de pri-miR394a.

A. Intensidades normalizadas de las bandas de digestión con Nucleasa S1. B. Intensidades normalizadas de las bandas de digestión con Nucleasa V1. C. Estructura de menor energía obtenida con el servidor Mfold con parámetros establecidos por defecto por los autores (37°C, NaCl 1M) (Zuker, 2003). D. Estructura alternativa obtenida incorporando restricciones derivadas de las determinaciones experimentales. Incluyendo estas restricciones se logra una mejor representación de los datos experimentales. Rojo: región doble hebra. Azul: región simple hebra. Flechas rojas: sitios donde ocurren los cortes secuenciales por parte de DCL1.

En este caso el tallo inferior doble hebra presenta una distancia de 19 nucleótidos hasta el extremo 3' del sitio del primer corte. Para este precursor las regiones definidas como importantes para el procesamiento son un tallo inferior y un tallo superior doble hebra (Bologna et al. 2013). Nuestro mapa estructural confirma la presencia de ambos determinantes. A diferencia de lo que obtuvimos para el pri-miR319a, este precursor no muestra una mayor estabilidad de la doble hebra en la región del dúplex miR394a/miR394a\*, por lo tanto no podemos confirmar que todos los casos de procesamiento secuencial impliquen la identificación del miARN por presentar una hibridación más fuerte.



Una observación general respecto a todos los precursores ensayados es que **la estructura tipo hélice A esperada para estas moléculas con grandes regiones apareadas no es homogénea**. El hecho de que se observen distintas intensidades en el ataque por V1 indica que la estabilidad de la hélice A es diferentes en las distintas regiones. Es interesante destacar que en todos los casos **los sitios donde se producen los cortes por DCL1 se encuentran en el contexto de regiones que presentan carácter tipo hélice A**. Al mismo tiempo, ninguno de los sitios se encuentra en las regiones donde la digestión por V1 es más agresiva. Esto podría estar reportando una necesidad de que la estabilidad de la hélice en esta región sea relativamente baja para permitir la acción de DCL1.

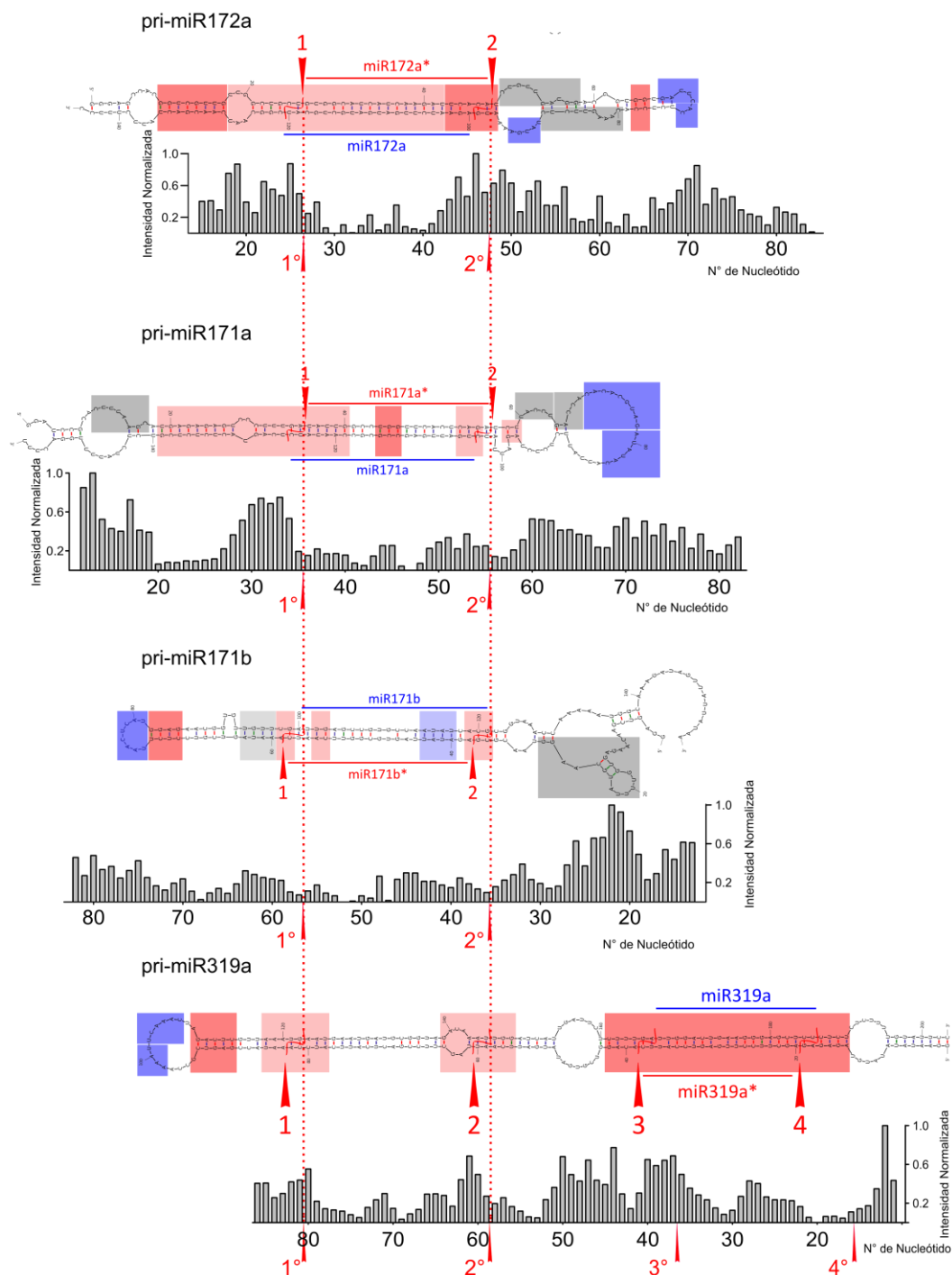
Por tratarse de una digestión enzimática, estos ensayos son muy dependientes de la conformación de regiones relativamente grandes del sustrato. Es decir, el sitio activo de las enzimas requiere un mínimo de nucleótidos en una cierta conformación para llevar a cabo su acción. Para independizarnos de este factor buscamos una técnica complementaria, la hidrólisis inducida por Terbio (III).

## 6.2. Mapeo de flexibilidad por hidrólisis inducida por Terbio (III)

Otra de las técnicas utilizadas con el fin de definir la estructura del ARN es la hidrólisis inducida por cationes trivalentes como el Terbio (III). Este tipo de iones posee características básicas en medio acuoso. Al unirse al el ARN da lugar a la desprotonación del 2'OH y lo activa para el ataque nucleofílico sobre el enlace fosfodiéster del esqueleto de la hebra de ARN. Este ataque es el mecanismo por el cual ocurre naturalmente la hidrólisis del ARN y es una de las causas de su relativa inestabilidad respecto al ADN. La acción del Terbio (III) acelera este mecanismo natural. Para que la hidrólisis se produzca es necesario que la cadena adquiera una conformación denominada “en línea” entre el átomo de Oxígeno del grupo 2'OH y el átomo de Fosforo del enlace fosfodiéster anterior. La probabilidad de adquirir esta conformación depende de la flexibilidad de la hebra.

Los resultados obtenidos se muestran en la Figura 6.8, donde la altura de las barras proviene de cuantificar la intensidad de las bandas radioactivas separadas en geles de poliacrilamida desnaturalizantes luego del tratamiento de la muestra de cada precursor (marcado en el extremo 5' con fosforo radioactivo) con Terbio (3+) 10 mM durante 15 minutos a temperatura ambiente. Los precursores que analizamos son los mismos que en la sección anterior, el del miR172a, miR171a, miR171b y miR319a.

Un análisis global de estos resultados nos muestra que al alinear los perfiles de hidrólisis según el sitio donde DCL1 realiza el primer corte observamos un cambio abrupto en la flexibilidad de la cadena. Podemos observar que **la región que antecede al sitio de corte presenta una flexibilidad alta, mientras que la región inmediatamente después del mismo sitio presenta una flexibilidad relativamente baja**. Este tipo de patrón podría ser determinante para el reconocimiento de este sitio por parte de DCL1 en conjunto con las proteínas accesorias. Como mencionamos anteriormente, el sitio del primer corte es el que consideramos de mayor importancia dado que es el que define el registro a partir del cual DCL1 desempeña su actividad tipo Dicer, efectuando cortes de forma procesiva liberando fragmentos de *ca.* 21 nucleótidos. **Otra particularidad que surge de los perfiles de flexibilidad que muestran las regiones donde ocurren el resto de los cortes consiste en que todos estos sitios se encuentran en regiones de flexibilidad relativamente alta**. Por lo tanto, podemos pensar que una hebra flexible es favorable para el procesamiento por parte de DCL1.



**Figura 6.8 Mapeo por hidrólisis inducida por Terbio (3+) de precursores de micro-ARNs**

En cada caso se muestra la estructura secundaria determinada en la sección anterior para cada precursor junto con el gráfico de la intensidad relativa de las bandas obtenidas luego de la hidrólisis inducida por Terbio (3+). Mayor intensidad corresponde a mayor hidrólisis que a su vez corresponde a mayor flexibilidad del esqueleto de la hebra. La cuantificación de las bandas se realizó empleando el software SAFA (Laederach et al., 2009). Los distintos precursores fueron alineados según el extremo 3' del sitio del primer corte de DCL1.

### 6.3. Discusión

Los mapas estructurales experimentales de los precursores de miARN resultan en una herramienta necesaria para la validación de las teorías sobre determinantes estructurales que surgen de plegamientos calculados *in silico*. Una de las limitaciones de las estructuras calculadas consiste en que no se tienen en cuenta la posibilidad de formación de pares de bases no canónicos, lo cual puede resultar confuso a la hora del análisis y búsqueda de estructuras particulares alrededor de los sitios de corte. La importancia de la formación de pares de base no canónicos en las moléculas de ARN en general se denota no sólo por la prevalencia de estos apareamientos en las estructuras conocidas sino también por su rol en la funcionalidad de estas moléculas (Lemieux & Major 2002; Dethoff et al. 2012).

A través del mapeo por digestión enzimática logramos validar la existencia de la región con estructura doble hebra de aproximadamente 17 nucleótidos necesaria para el correcto procesamiento de los precursores ensayados. En conjunto con los datos de flexibilidad de la hebra podemos describir esta región con algo más de profundidad. Podemos decir que la estabilidad de la hélice no es homogénea a lo largo de todo el tallo de 17 nt, siendo la región distal respecto del corte la que presenta un carácter de hélice A más pronunciado, mientras que el sitio de corte se encuentra en el contexto de regiones con un carácter de hélice relativamente débil. De acuerdo con estas observaciones, la flexibilidad se ve aumentada en la región cercana al sitio del primer corte y el perfil cambia abruptamente a una región de menor flexibilidad. Estos detalles estructurales podrían ser necesarios para permitir la interacción con el complejo de procesamiento y acción de la enzima.

En relación a la interacción de los precursores con la enzima consideramos que nuestros datos experimentales suman evidencia al modelo que planteamos en la sección 1.7 de la introducción de este trabajo, centrado en la importancia de la longitud del tallo que precede al sitio del primer corte y a lo que conocemos sobre la interacción con el sustrato de los dominios ARNasa III junto con los dsRBDs.

## 7. Conclusiones finales

Durante este trabajo encontramos que el primer dsRBD de DCL1, DCL1-A, se encuentra mayormente desestructurado en su forma libre. El ensamble conformacional de DCL1-A revela que el extremo C-terminal presenta una tendencia a explorar estructuras de hélice alfa. Esta estructura hace que uno de los sitios de interacción con el ARNdh se encuentre preformado. La combinación de la alta flexibilidad de la cadena con la presencia de uno de los sitios de reconocimiento del sustrato preformado sugiere que muy probablemente esta proteína encuentre a su ligando a través de un mecanismo tipo “pesca con mosca” (Shoemaker et al. 2000). Este mecanismo implicaría una ventaja para la enzima por acelerar la velocidad de encuentro con el ARN blanco.

DCL1-A adquiere un plegamiento tipo dsRBD al interactuar con el ARNdh pero se mantiene desplegado al interactuar con ADN. En el contexto de la enzima completa esta diferencia podría estar induciendo cambios conformacionales que permitan identificar de forma específica al sustrato, diferenciándolo de otros ácidos nucleicos.

A partir de la estructura adquirida por el dominio podemos inferir que la forma plegada de DCL1-A interacciona con el ARNdh a través de la mayoría de los sitios de reconocimiento canónicos para los dsRBDs. Esto nos lleva a considerar que probablemente al plegarse esté cumpliendo funciones similares a los dsRBDs de otras ARNasas III. Como expresamos en la sección 1.7 de la introducción, por la evidencia disponible podemos pensar que los dominios ARNasa III en tándem con un dsRBD escinden el ARNdh en fragmentos de una longitud que depende de la región que ocupa el arreglo de estos dominios sobre el sustrato. En el caso de Dicer de humanos esta longitud es de 15 nucleótidos, en lugar de los 21 que libera cuando está completa (Ma et al. 2012). Por otro lado, determinamos experimentalmente que la longitud de ARN en conformación de hélice A que precede al sitio del primer corte es de aproximadamente 17 pb. Considerando estas evidencias en conjunto es razonable concluir que la región cubierta por los dominios ARNasa III en tándem con DCL1-A de DCL1 sea probablemente de *ca.* 17 pb y muy probablemente sea esta forma de interacción con el sustrato la que resulte en la definición del sitio del primer corte.

Sabemos que el segundo dsRBD de DCL1, DCL1-B, es necesario para la localización de la enzima en los cuerpos D dentro del núcleo, y que este dominio es capaz de localizar en el núcleo mientras que DCL1-A no lo hace. Por otro lado existe evidencia de que una enzima Dicer de humanos a la cual se agregó una segunda copia del dsRBD C-terminal es capaz de ingresar al núcleo con mayor eficiencia que la enzima silvestre (Doyle et al. 2013). Esto nos lleva a pensar en un posible rol para DCL1-A confiriendo un conector altamente flexible entre DCL1-B y el resto de la enzima, permitiendo la disponibilidad de la señal de localización subcelular de la enzima.

En conjunto consideramos que probablemente la plasticidad estructural de DCL1-A cumpla roles de importancia para las distintas capacidades de DCL1, tanto a nivel de localización subcelular, como en la eficiencia y especificidad de la reacción catalizada.

## 8. APENDICE

### 8.1. Apéndice I

Código de Python para correr la rutina de generación de un ensamble conformacional con PyRosetta:

```
#!/usr/bin/python
import time
import os, sys
import numpy as np
from matplotlib import pyplot as pl
import rosetta as ro

def randomMove(sco):
    '''
    La funcion produce un cambio de Phi y Psi
    en un residuo al azar, y evalua si mejora
    o empeora el score de la estructura
    El argumento es el Score de la estructura actual
    Para simplificar el bucle while del programa, devuelve
    False (salir del bucle) si el cambio es favorable
    y True (seguir en el bucle) si es desfavorable
    '''
    global NewSco
    #cambia el phi/psi de un residuo al azar
    #y guarda los valores originales de Phi y Psi
    #por si el cambio es desfavorable
    randRes=np.random.randint(1,NumResidues+1)
    initialPhi=pose.phi(randRes)
    initialPsi=pose.psi(randRes)
    #Usa la informacion de restriccion por secondary shift
    #si la encuentra
    if len(CompletePhi[randRes])>0:
        numeroAlAzar=np.random.randint(len(CompletePhi[randRes]))
        Phi=float(CompletePhi[randRes][numeroAlAzar])
        Psi=float(CompletePsi[randRes][numeroAlAzar])
    #si no, toma un valor de la base de datos de FM al azar
    else:
        resi=seq[randRes-1]
        #print "%s%i sin datos" %(resi,r)
        choice=np.random.randint(len(FMpsi[resi]))
        Phi=FMphi[resi][choice]
        Psi=FMpsi[resi][choice]
    #cambia los angulos Phi y Psi de la estructura
    #y calcula el score resultante
    pose.set_phi(randRes,Phi)
    pose.set_psi(randRes,Psi)
    NewSco=scoreFun(pose)
    if NewSco < sco:
        #si el score mejora, acepta la estructura
        return False
    else:
        #Para permitir salir de pozos de potencial chicos
        #da una probabilidad de que se acepte la estructura
        #aunque el score sea mayor
        #La probabilidad disminuye como
        #p=exp(deltaScore/kT)
        #poniendo kT= 1 unidad de Score de Rosetta
        delta=sco-NewSco
        prob=np.exp(delta)
        azar=np.random.random()

        if azar > prob:
            #si no acepta la estructura, vuelve a los valores
            #originales de Phi y Psi
            pose.set_phi(randRes,initialPhi)
```

```

        pose.set_psi(randRes,initialPsi)
        return True
    else:
        #print "METROPOLIS!!! prob=%f; azar=%f" % (prob,azar)
        return False

#Diccionario para traducir nombres de residuo de tres a una letra

three2one={
"ALA":"A",
"ARG":"R",
"ASN":"N",
"ASP":"D",
"CYS":"C",
"GLN":"Q",
"GLU":"E",
"GLY":"G",
"HIS":"H",
"ILE":"I",
"LEU":"L",
"LYS":"K",
"MET":"M",
"PHE":"F",
"PRO":"P",
"SER":"S",
"THR":"T",
"VAL":"V",
"TRP":"W",
"TYR":"Y"
}

Zero=time.time()

try:
    info=open("%s/info.txt"%sys.argv[1],'r').readlines()
except:
    print """
Uso:
%s carpeta_para_datos

Calcula conformeros usando rosetta
la carpeta debe contener el file info.txt

    """ % sys.argv[0]
    sys.exit()

#Lee el file con info y crea un diccionario con los datos necesarios
#info.txt tiene el siguiente formato:
#
#sequence:          GHMMHPVRELQERCQQQAEGLEYKASRSGNTATVEVFIDGVQVGVAQNPQKKMAQKLAARNALALKEKEI
#AESKEKHINNGN
#FMdatabase:        /Users/admin/Documents/flexiblemeccano/database/database_options/phi_ps
#i.txt
#phipsi1:           /Users/admin/Dropbox/Irina/pyRosetta/0_0_helix_10k/Analysis/AngulosAll
#phipsi2:           /Users/admin/Dropbox/Irina/pyRosetta/0_0_helix_10k/Analysis/AngulosAll
#...
#Conformers:        10000
#Begin Conf:         1000
#baseName:           nb22
#
#Las lineas comentadas con '#' no se leen

salida=open("%s/out.dat"%sys.argv[1],'w')
infoData=dict()
for l in info:
    if not "#" in l:
        param, dato = l.split(":")
        infoData[param.strip()]=dato.strip()
#~ print infoData
protSequence=infoData["sequence"]
FMdatabase=infoData["FMdatabase"]

```

```

filesConAngulos=[infoData[x] for x in infoData.keys() if "phipsi" in x]
AngNames=[x.split('/')[0] for x in filesConAngulos]
TotalConformers=int(infoData["Conformers"])
BeginConformer=int(infoData["Begin Conf"])
baseName=infoData["baseName"]

#Inicializa rosetta y genera la estructura de base
#La estructura esta definida como centroide
#tiene un pseudoatomo en lugar de la sidechain
ro.init()
pose=ro.pose_from_sequence(protSequence,"centroid")
seq=pose.sequence()
NumResidues=pose.total_residue()

#El objeto PyMOL_Mover permite visualizar las estructuras generadas en pymol
#Para usarlo:
#+ iniciar Pymol
#+ buscar el script PyMOLPyRosettaServer.py en la instalacion de PyRosetta
#+ Correr el siguiente comando en la linea de comando de pymol
#run PyMOLPyRosettaServer.py
#Esto deja abierto un canal para enviar datos a pymol
pymol=PyMOL_Mover()

#Usamos como funcion de score la funcion por defecto de Rosetta para
#estructuras con centroides en vez de full atom sidechains
scoreFun=get_cen_scorefxn()

#Busca el file con la base de datos de Flexible Meccano
try:
    FMdbase=open(FMdatabase,'r').readlines()
except:
    print """
    no se encontro la base de datos de Flexible Meccano en %s
    """ % infoData["FMdatabase"]
    sys.exit()

#La base de datos de FM tiene varios tipos de residuos
#que no son los aminoasidos estandar
residueTypes=set([x.split()[0] for x in FMdbase])

#Crea un diccionario con los valores de phi y psi
#de la base de datos de Flexible Meccano para cada tipo de residuo
FMphi=dict()
FMpsi=dict()
for r in residueTypes:
    if r in three2one.keys():#toma solo los residuos estandar
        print "leyendo base de datos FM para %s" % r
        FMphi[three2one[r]]=float(x.split()[1]) for x in FMdbase if len(x.split())==3 and x[:3]==r
        FMpsi[three2one[r]]=float(x.split()[2]) for x in FMdbase if len(x.split())==3 and x[:3]==r

#lee los datos de angulos fijados para distintos residuos
#se pueden usar tantos files como se quiera
#crea un diccionario para cada Phi y para Psi
#para cada file de datos y llena los diccionarios
#con los datos leidos en los files correspondientes

todosPhi=[dict() for x in range(len(filesConAngulos))]
todosPsi=[dict() for x in range(len(filesConAngulos))]
for i,ang in enumerate(filesConAngulos):
    print "leyendo %s"%ang
    angulos=open(ang,'r').readlines()
    todosPhi[i]=dict([int(x.split()[0]),x.split()[2:] for x in angulos if x.split()[1]== "Phi" and len(x.split())>2])
    todosPsi[i]=dict([int(x.split()[0]),x.split()[2:] for x in angulos if x.split()[1]== "Psi" and len(x.split())>2])

for i,j in enumerate(todosPhi):print AngNames[i],j.keys()

```

```

#crea un diccionario nuevo donde va a almacenar
#los datos de cada uno de los diccionarios de angulos fijos anteriores
#esto es necesario porque puede haber residuos presentes
#en un diccionario y ausentes en otro
#(NO incluye los datos de FM!!!)

CompletePhi=dict(zip(range(1,NumResidues+1), [[] for x in range(1,NumResidues+1)]))
CompletePsi=dict(zip(range(1,NumResidues+1), [[] for x in range(1,NumResidues+1)]))

#para cada residuo, si hay datos,
#extiende el diccionario correspondiente
for k in range(1,NumResidues+1):
    testPresent=[False for x in AngNames]
    for i,phd,psd in zip(range(len(AngNames)),todosPhi,todosPsi):
        try:
            CompletePhi[k].extend(phd[k])
            CompletePsi[k].extend(psd[k])
            testPresent[i]=True
        except:
            pass

    if True in testPresent:
        Tiene=[x for x in AngNames if testPresent[AngNames.index(x)]]
        print "Residuo %s%i con %s" % (seq[k-1],k, " ".join(Tiene))
        salida.write("Residuo %s%i con %s\n" % (seq[k-1],k, " ".join(Tiene)))
    else:
        print "Residuo %s%i SIN DATOS!" % (seq[k-1],k)
        salida.write("Residuo %s%i SIN DATOS!\n" % (seq[k-1],k))

#verifica el numero de phis y psis para cada residuo
#deberia haber igual numero

for k in CompletePhi.keys():
    print "Residuo %i, %i Phis y %i Psis" %(k, len(CompletePhi[k]), len(CompletePsi[k]))
    if len(CompletePhi[k]) != len(CompletePsi[k]):
        print "\n\n++++++\nPROBLEMA CON LAS RESTRICCIONES DE PHI/PSI!!!\n+++++\n"
        sys.exit()
    salida.write("Residuo %i, %i Phis y %i Psis\n" %(k, len(CompletePhi[k]), len(CompletePsi[k])))

#Crea arrays para guardar los valores de phi y psi
#de cada estructura generada
#para graficar los ramachandran al final
ToDrawPhis=np.zeros(NumResidues)
ToDrawPsis=np.zeros(NumResidues)

# va a probar como maximo 200 cambios de phi/psi al azar (vueltas)
# en cada vuelta va a hacer como maximo 100 pruebas
# de cambiar un phi/psi al azar hasta que baje la energia
vueltas=200
maxPruebas=100

#En el array scoreDraw se guardan los scores por vuelta para graficar al final
scoreDraw=np.zeros(vueltas+1)
previousTime=time.time()

#crea tantos conformeros como pide en el file info
for s in range(BeginConformer, TotalConformers+BeginConformer):
    decoysDone=0
    #en scoreEvolution se guarda la evolucion del score con las vueltas
    scoreEvolution=np.array([])

    #primero crea un conformero fijando valores de
    #phi y psi para cada residuo a lo largo de toda la secuencia
    for r in range(1,NumResidues+1):
        #Si tiene datos fijados para el residuo, toma un par al azar
        if len(CompletePhi[r])>0:
            numeroAlAzar=np.random.randint(len(CompletePhi[r]))
            Phi=float(CompletePhi[r][numeroAlAzar])

```



```

        Psi=float(CompletePsi[r][numeroAlAzar])

    #si no toma un valor de la base de datos de FM al azar
    else:
        resi=seq[r-1]
        #print "%s%i sin datos" %(resi,r)
        choice=np.random.randint(len(FMpsi[resi]))
        Phi=FMphi[resi][choice]
        Psi=FMpsi[resi][choice]

    pose.set_phi(r,Phi)
    pose.set_psi(r,Psi)
    #calcula el score y lo guarda en el vector de scoreEvolution
    sco=scoreFun(pose)
    scoreEvolution=np.append(scoreEvolution,sco)

    #Prueba *vueltas* veces cambiar los phiPsi
    for i in range(vueltas):
        #hace hasta *pruebas* pruebas de cambiar un angulo al azar
        #para mejorar el score
        pruebas=0
        muchasPruebas=False
        while randomMove(sco): #randomMove devuelve False cuando mejora el score y sale
del loop
            pruebas+=1
            if pruebas>maxPruebas:
                print "Muchas pruebas!!"
                muchasPruebas=True
                break
            if muchasPruebas:break
        #permite ver en pymol como se arma la estructura generada
        #VUELVE TODO MUUY LENTO!!!!
        #~ pymol.apply(pose)

        sco=NewSco
        scoreEvolution=np.append(scoreEvolution,sco)

    #permite ver en pymol cada estructura generada
    #VUELVE TODO MUUY LENTO!!!!
    #~ pymol.apply(pose)

    #Guarda el decoy y elimina las lineas con los centroides
    pose.dump_pdb("%s/nb22_%04i_CEN.pdb"%(sys.argv[1],s))
    CenPDB=open("%s/nb22_%04i_CEN.pdb"%(sys.argv[1],s),"r")
    NoCenPDB=open("%s/ia_83.pdb"%(sys.argv[1],s),"w")
    CP=CenPDB.readlines()
    for l in CP:
        if not "CEN" in l:
            NoCenPDB.write(l)
    NoCenPDB.close()
    CenPDB.close()
    os.remove("%s/nb22_%04i_CEN.pdb"%(sys.argv[1],s))

    print "Decoy % i, vueltas %i, score %f, %f segundos" % (s, i, sco, (time.time()-
previousTime))
    salida.write("nb22_%04i.pdb, score %6.2f, %i pruebas, %f segundos\n"%\
(s,sco, i, (time.time()-previousTime)))
    previousTime=time.time()
    if scoreEvolution.shape[0]<vueltas:
        #si se cortaron las vueltas por muchas pruebas, completa
        #el vector scoreEvolution con el ultimo valor para
        #poder graficarlo
        agregar=np.array([scoreEvolution[-1] for x in range(vueltas-
scoreEvolution.shape[0]+1)])
        scoreEvolution=np.append(scoreEvolution,agregar)
        #Agrega los phis y psis del nuevo conformero a los datos a graficar
        ToDrawPhis=np.vstack([ToDrawPhis,np.array([pose.phi(x) for x in range(1,NumResidues+
1)])])
        ToDrawPsis=np.vstack([ToDrawPsis,np.array([pose.psi(x) for x in range(1,NumResidues+
1)])])
        scoreDraw=np.vstack([scoreDraw,scoreEvolution])

```

```

#Dibuja las evoluciones de los scores de todos los conformeros generados
pl.plot(range(scoreDraw.shape[1]),scoreDraw.T)
pl.xlabel("iteration")
pl.ylabel("energy")
pl.title("Evolucion scores %s"%sys.argv[1])
pl.savefig("%s/energies.pdf"%(sys.argv[1]))
pl.clf()

#Dibuja el Ramachandran de los decoys generados
#para cada residuo
for i,res in enumerate(seq):
    pl.plot(ToDrawPhis[:,i],ToDrawPsis[:,i], 'ko',alpha=0.5)
    pl.title("%s%i"%(res,i+1))
    pl.xlim(-180,180)
    pl.ylim(-180,180)
    pl.xlabel("Phi")
    pl.ylabel("Psi")
    pl.savefig("%s/Rama_%03i%s.pdf"%(sys.argv[1],i+1,res))
    pl.clf()
    print "ploteado %s%i" % (res, i+1)

#Dibuja el contenido de estructura secundaria promedio
#por residuo
allAlphaR=[]
allBetaS=[]
allPPII=[]
allAlphaL=[]
for i,res in enumerate(seq):
    PhisPsis=zip(ToDrawPhis[:,i],ToDrawPsis[:,i])
    #las definiciones de las regiones estan hechas de acuerdo a
    #Ozenne et al., JACS 2012, Figure 2
    alphaR=[1 for x in PhisPsis if x[0]<0 and x[1]>-120 and x[1]<50]
    betaS=[1 for x in PhisPsis if x[0]<-100 and (x[1]>=50 or x[1]<=-120)]
    PPII=[1 for x in PhisPsis if x[0]<0 and x[0]>-100 and (x[1]>=50 or x[1]<=-120)]
    alphaL=[1 for x in PhisPsis if x[0]>0]
    allAlphaR.append(len(alphaR))
    allBetaS.append(len(betaS))
    allPPII.append(len(PPII))
    allAlphaL.append(len(alphaL))
allAlphaR=np.array(allAlphaR)/float(TotalConformers)
allBetaS=np.array(allBetaS)/float(TotalConformers)
allPPII=np.array(allPPII)/float(TotalConformers)
allAlphaL=np.array(allAlphaL)/float(TotalConformers)
fractionData=np.column_stack((allAlphaR,allBetaS,allPPII,allAlphaL))
np.savetxt("%s/SecStructFractions.csv"%sys.argv[1], fractionData, delimiter=",")
residues=np.arange(1,NumResidues+1)
f, (ax1, ax2, ax3, ax4) = pl.subplots(4, sharex=True, sharey=True)
ax1.plot(residues,allAlphaR,label="alpha")
ax1.set_ylabel("alpha")
ax2.plot(residues,allBetaS,label="beta")
ax2.set_ylabel("beta")
ax3.plot(residues,allPPII,label="PPII")
ax3.set_ylabel("PPII")
ax4.plot(residues,allAlphaL,label="alpha-L")
ax4.set_ylabel("alpha-L")
ax4.set_xlabel("residue")
ax1.set_title("Sec. struct fractions")
f.subplots_adjust(hspace=0)
pl.savefig("%s/SesStructFractions.pdf"%(sys.argv[1]))

#Grafica los valores de phi/psi de la base de datos de FM
#for k in FMphi.keys():
#    pl.plot(FMphi[k],FMpsi[k], 'ko',alpha=0.5)
#    pl.ylim(-180,180)
#    pl.xlim(-180,180)
#    pl.ylabel("Psi")
#    pl.xlabel("Phi")
#    pl.title("FM database for %s, %i values"%(k,len(FMphi[k])))
#    pl.savefig("FM_%s_Rama.pdf"%k)
#    pl.clf()

```

```
#
#sys.exit()
```

## 8.2. Apéndice II

Código de Python del algoritmo genético que ajusta la población de confórmers del ensamble de DCL1-A:

```
#!/usr/bin/python
import os, sys
import numpy as np
from matplotlib import pyplot as pl
import random

#creamos la clase de objeto cromosoma y le damos atributos
class cromosoma:
    '''
    El cromosoma contiene una lista de numeros (datos)
    una fitness respecto de un target
    una funcion de chequeo de fitness
    y una funcion de crossover
    '''
    def __init__(self, inData):
        self.datos=inData
        #los datos tienen la forma
        #[[SpartaFile1,array([ca1,ca2,...],[cb1,cb2,...],[col,co2,...])],[SpartaFile2],
        array([ca1,ca2,...],...)]
        self.mide()
        self.promedio()
        self.leeNombres()
        self.fitness=-1
    def checkFitness(self,target):
        '''
        El fitness es la diferencia entre el promedio de los numeros del cromosoma y el
        target al cuadrado
        '''
        self.fitness=np.sum((self.total-target)**2)
        return self.fitness
    def crossover(self,cromosoma2):
        '''
        En un crossover, un cromosoma intercambia los datos con otro a partir de un punt
        o aleatorio de su secuencia
        '''
        if len(self.datos)!=len(cromosoma2.datos):
            return "different lengths!"
        point=np.random.randint(len(self.datos))
        buff1=self.datos[:]
        buff2=cromosoma2.datos[:]
        self.datos[0:point]=cromosoma2.datos[0:point]
        cromosoma2.datos[0:point]=buff1[0:point]
        self.leeNombres()
        cromosoma2.leeNombres()
        self.promedio()
        cromosoma2.promedio()
        return point, self.nombres, cromosoma2.nombres
    def promedio(self): ##promedio de los Sec Shifts dentro del cromosoma
        todosCS=self.datos[0][1]
        for i in self.datos[1:]:
            todosCS=np.sum((todosCS,i[1]),axis=0)
        self.total=todosCS/self.largo
        return self.total
    def leeNombres(self):
        self.nombres=[x[0].split("/")[-1] for x in self.datos]
        return self.nombres
    def mide(self):
        self.largo=len(self.datos)
        return self.largo
    def mutar(self, pool):
```

```

        posicion=random.randint(self.largo)
        genAlAzar=random.randint(len(pool))
        sacado=self.datos[posicion]
        metodo=pool[genAlAzar]
        self.datos[posicion]=metodo
        return posicion, metodo, sacado

#definimos la funcion de entrecruzamiento entre cromosomas
def nextGeneration(cromosomaList):
    '''
    En una generacion se toman los cromosomas de a pares
    segun su orden en la lista de ingreso
    y se los cruza
    '''
    if len(cromosomaList)%2 != 0: return "Tienen que ser pares!!"
    for i in range(len(cromosomaList)/2):
        cromosomaList[2*i].crossover(cromosomaList[2*i+1])

Dir=sys.argv[1]
OutDir="AjusteGenAl1007"

#los archivos de output se guardan en la carpeta OutDir
try:
    os.stat("%s/%s/" % (Dir, OutDir))
except:
    os.mkdir("%s/%s/" % (Dir, OutDir))

#Crea listas con los nombres de archivos de los outputs de Sparta
spartaDir="%s/Output/SPARTA/" %Dir
SpartaFiles = [ spartaDir+ f for f in os.listdir(spartaDir)\
    if os.path.isfile(os.path.join(spartaDir,f))\
    and 'a_83' in f]

SpartaFiles.sort()

print '''
+++++
+ Hay %i estructuras!!!
+++++
''' % len(SpartaFiles)

exptCAfile=open('/home/diegog/flexiblemeccano/simulations/150806_Irina/Expt_278K/filttdca
.ca', 'r').readlines()
exptCBfile=open('/home/diegog/flexiblemeccano/simulations/150806_Irina/Expt_278K/filttdcb
.cb', 'r').readlines()
exptCOfile=open('/home/diegog/flexiblemeccano/simulations/150806_Irina/Expt_278K/filttdco
2.co', 'r').readlines()

#Crea una lista con los numeros de residuo y la secuencia.
resnums=range(1,83)
sequence="GHMMHPVRELQERCQQQAEGLEYKASRSGNTATVEVFIDGVQVGVAQNPQKKMAQKLAARNALAAALKEKEIAESKEKH
INNGN"

#Crea diccionarios tipo {numResiduo:SecShift,...} para CA, CB y CO
arrayExpCA=np.zeros(84)
arrayExpCB=np.zeros(84)
arrayExpCO=np.zeros(84)
exptCA=dict([int(x.split()[0]),float(x.split()[1])] for x in exptCAfile if not "#" in x)
exptCB=dict([int(x.split()[0]),float(x.split()[1])] for x in exptCBfile if not "#" in x)
exptCO=dict([int(x.split()[0]),float(x.split()[1])] for x in exptCOfile if not "#" in x)
resConCA=np.array(exptCA.keys())
resConCB=np.array(exptCB.keys())
resConCO=np.array(exptCO.keys())
for k in exptCA.keys():
    arrayExpCA[k]=exptCA[k]
for k in exptCB.keys():
    arrayExpCB[k]=exptCB[k]
for k in exptCO.keys():
    arrayExpCO[k]=exptCO[k]
TodosExpSS= np.vstack((arrayExpCA,arrayExpCB,arrayExpCO))

```

```

##arma arrays de ceros para cada nucleo, luego pone los valores p los q hay valores experimentales,
##los residuos p los q no hay valor experimental quedan con valor cero, o sea dejamos afuera el valor calculado
strucAll=dict()
arrayCA=np.zeros(84)
arrayCB=np.zeros(84)
arrayCO=np.zeros(84)

#Lee todos los archivos de SPARTA
for i,sparta in enumerate(SpartaFiles):
    sp=open(sparta,'r').readlines()
    #Para cada CS crea un diccionario de estructura {spartaFile:array[secondaryShifts],..}
    ..)
    losCaPresentes=[[int(x.split()[0]), float(x.split()[3])] for x in sp[28:] if x.split()[2]=="CA" and int(x.split()[0]) in resConCA]
    losCbPresentes=[[int(x.split()[0]), float(x.split()[3])] for x in sp[28:] if x.split()[2]=="CB" and int(x.split()[0]) in resConCB]
    losCoPresentes=[[int(x.split()[0]), float(x.split()[3])] for x in sp[28:] if x.split()[2]=="C" and int(x.split()[0]) in resConCO]
    for i,cs in losCaPresentes:
        arrayCA[i]=cs
    for i,cs in losCbPresentes:
        arrayCB[i]=cs
    for i,cs in losCoPresentes:
        arrayCO[i]=cs
    strucAll[sparta]=np.vstack((arrayCA,arrayCB, arrayCO))
    if i%100 ==0: print "procesado hasta %s" % sparta

#Definimos cromosomas al azar
totalGenes=len(listStrucAll)
print "GENES",totalGenes
largoCrom = 100 #tamano de los cromosomas, tiene que ser fraccion entera del total de genes
if totalGenes%largoCrom !=0:
    print "El tamano de los cromosomas tiene que ser fraccion entera del total de genes !!!"
    #sys.exit()
totalCrom=totalGenes/largoCrom
repeticiones=10
CopyListStrucAll=listStrucAll[:]
cromosomas=[]

##popea estructuras al azar de la lista de los nombres de las estructuras y los pone en el nuevo cromosoma
for r in range(repeticiones):
    print "repeticion %i" %r
    for crom in range(totalCrom):
        nuevoCrom=[]
        for g in range(largoCrom):
            nuevoCrom.append(CopyListStrucAll.pop(np.random.randint(len(CopyListStrucAll))))
        nuevoCrom=cromosoma(nuevoCrom)
        cromosomas.append(nuevoCrom)
    CopyListStrucAll=listStrucAll[:]

#creo dos listas para seguir el fitness del mejor gen
#y el fitness promedio de los cromosomas
#a medida que pasan las generaciones
fitnessEvolution=[]
averageEvolution=[]
target=TodosExpSS

#cada vuelta es una generacion
Vueltas=1000 ##numero de vueltas de mezcla de cromosomas
for vuelta in range(Vueltas):
    print "Vuelta %i" %vuelta
    #creo una lista con los fitness de cada cromosoma
    #la pego a la lista de cromosomas para ordenarlos
    #segun su fitness

```

```

cromosomasFitness=[x.checkFitness(target) for x in cromosomas]
PresortedCroms=zip(cromosomasFitness,cromosomas)
PresortedCroms.sort()
#mezclamos el orden de los cromosomas dentro de bloques de a diez
#p aumentar la probabilidad de que suban estructuras que quedaron abajo
sortedCroms=[]
sizeBloques=10
bloques=totalCrom*repeticiones/sizeBloques
for i in range(bloques):
    block=PresortedCroms[i*sizeBloques:(i+1)*sizeBloques]
    random.shuffle(block)
    sortedCroms.extend(block)

#incorporo a las listas de seguimiento la fitness del mejor cromosoma
#y la fitness promedio de todos los cromosomas
fitnessEvolution.append(PresortedCroms[0][0])
averageEvolution.append(sum([x[0] for x in PresortedCroms])/float(len(PresortedCroms
)))

#Para hacer el crossover tomo los cromosomas ordenados y los pongo en la lista toCros
ss
toCross=[x[1] for x in sortedCroms]
nextGeneration(toCross)
#para la proxima generacion, los cromosomas son los ya cruzados
cromosomas=toCross

if vuelta%100==0:
    pl.bar(np.arange(84),TodosExpSS[0],color='k', alpha=0.2, align='center', label="
Experimentales")
    pl.plot(np.arange(84),PresortedCroms[0][1].promedio()[0,:], 'r-
', label="best cromosoma")
    pl.title("CA vuelta %i" % vuelta)
    pl.legend(loc=0)
    pl.savefig("%s/%s/CASecShiftBestCrom%sCroms%sloci%sVue_v%i.pdf" %(Dir, OutDir, t
otalCrom*repeticiones, largoCrom, Vueltas,vuelta))
    pl.clf()

    pl.bar(np.arange(84),TodosExpSS[1],color='k', alpha=0.2, align='center', label="
Experimentales")
    pl.plot(np.arange(84),PresortedCroms[0][1].promedio()[1,:], 'r-
', label="best cromosoma")
    pl.title("CB vuelta %i" % vuelta)
    pl.legend(loc=0)
    pl.savefig("%s/%s/CBSecShiftBestCrom%sCroms%sloci%sVue_v%i.pdf" %(Dir, OutDir, t
otalCrom*repeticiones, largoCrom, Vueltas,vuelta))
    pl.clf()

    pl.bar(np.arange(84),TodosExpSS[2],color='k', alpha=0.2, align='center', label="
Experimentales")
    pl.plot(np.arange(84),PresortedCroms[0][1].promedio()[2,:], 'r-
', label="best cromosoma")
    pl.title("CO vuelta %i" % vuelta)
    pl.legend(loc=0)
    pl.savefig("%s/%s/COSecShiftBestCrom%sCroms%sloci%sVue_v%i.pdf" %(Dir, OutDir, t
otalCrom*repeticiones, largoCrom, Vueltas,vuelta))
    pl.clf()

sortedCroms.sort()

outFile=open("%s/%s/FitnessEvol%sCroms%sloci%sVue.out" %(Dir, OutDir, totalCrom*repetici
ones, largoCrom, Vueltas),"w")
print "fitness finales"
outFile.write("fitness finales\n")

#Hago un plot con la evolucion de los fitness al paso de las generaciones
pl.plot(fitnessEvolution, label="best fit")
pl.title("%s cromosomas de %s conformeros" %(totalCrom*repeticiones, largoCrom))
pl.xlabel("generacion")
pl.ylabel("fitness")
pl.legend(loc=0)
pl.savefig("%s/%s/FitnessEvol%sCroms%sloci%sVue.pdf" %(Dir, OutDir, totalCrom*repeticion

```

```

es, largoCrom, Vueltas))
pl.show()
pl.clf()

#Grafico los Sec Shifts calculados para el mejor cromosoma vs los experimentales
pl.bar(np.arange(84),TodosExpSS[0],color='k', alpha=0.2, align='center', label="Experimentales")
pl.plot(np.arange(84),sortedCroms[0][1].promedio()[0,:], 'r-', label="best cromosoma")
pl.title("CA vueltas %i" % Vueltas)
pl.legend(loc=0)
pl.savefig("%s/%s/CASecShiftBestCrom%sCroms%sloci%sVue.pdf" % (Dir, OutDir, totalCrom*repeticiones, largoCrom, Vueltas))
pl.show()
pl.clf()

pl.bar(np.arange(84),TodosExpSS[1],color='k', alpha=0.2, align='center', label="Experimentales")
pl.plot(np.arange(84),sortedCroms[0][1].promedio()[1,:], 'r-', label="best cromosoma")
pl.title("CB vueltas %i" % Vueltas)
pl.legend(loc=0)
pl.savefig("%s/%s/CBSecShiftBestCrom%sCroms%sloci%sVue.pdf" % (Dir, OutDir, totalCrom*repeticiones, largoCrom, Vueltas))
pl.show()
pl.clf()

pl.bar(np.arange(84),TodosExpSS[2],color='k', alpha=0.2, align='center', label="Experimentales")
pl.plot(np.arange(84),sortedCroms[0][1].promedio()[2,:], 'r-', label="best cromosoma")
pl.title("CO vueltas %i" % Vueltas)
pl.legend(loc=0)
pl.savefig("%s/%s/COSecShiftBestCrom%sCroms%sloci%sVue.pdf" % (Dir, OutDir, totalCrom*repeticiones, largoCrom, Vueltas))
pl.show()
pl.clf()

```

## Bibliografía

- Acevedo, R. et al., 2015. Helical Defects in MicroRNA Influence Protein Binding by TAR RNA Binding Protein. *Plos One*, 10(1), p.e0116749. Available at: <http://dx.plos.org/10.1371/journal.pone.0116749>.
- Addo-Quaye, C. et al., 2009. Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA (New York, N.Y.)*, 15(12), pp.2112–2121.
- Allen, E. & Howell, M.D., 2010. miRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Seminars in cell & developmental biology*, 21(8), pp.798–804. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20359543> [Accessed September 10, 2015].
- Aukerman, M.J. & Sakai, H., 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *The Plant cell*, 15(11), pp.2730–41. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=280575&tool=pmcentrez&rendertype=abstract> [Accessed April 16, 2015].
- Auyeung, V.C. et al., 2013. Beyond secondary structure: Primary-sequence determinants license Pri-miRNA hairpins for processing. *Cell*, 152(4), pp.844–858. Available at: <http://dx.doi.org/10.1016/j.cell.2013.01.031>.
- Barraud, P. et al., 2011. An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *The EMBO Journal*, 30(20), pp.4223–4235.
- Bass, B.L., Hurst, S.R. & Singer, J.D., 1994. Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Current biology : CB*, 4(4), pp.301–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7922339> [Accessed November 6, 2015].
- Bax, A. & Grishaev, A., 2005. Weak alignment NMR: A hawk-eyed view of biomolecular structure. *Current Opinion in Structural Biology*, 15(5), pp.563–570.
- Bernadó, P. et al., 2005. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 102(1), pp.17002–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1287987&tool=pmcentrez&rendertype=abstract>.
- Bertoncini, C.W. et al., 2005. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5), pp.1430–1435.
- Bologna, N.G. et al., 2009. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *The EMBO journal*, 28(23), pp.3646–3656.
- Bologna, N.G. et al., 2013. Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Research*, 23(10), pp.1675–1689.



- Bologna, N.G. & Voinnet, O., 2014a. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annual review of plant biology*, 65, pp.473–503. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24579988> [Accessed June 22, 2015].
- Bologna, N.G. & Voinnet, O., 2014b. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annual review of plant biology*, 65(June 2015), pp.473–503. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24579988>.
- Borges, A., Tauler, R. & Juan, A. de, 2005. Application of multivariate curve resolution to the temperature-induced unfolding of  $\alpha$ -chymotrypsin. *Analytica Chimica Acta*, 544(1-2), pp.159–166. Available at: <http://www.sciencedirect.com/science/article/pii/S0003267005001637> [Accessed August 24, 2015].
- Borsani, O. et al., 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, 123(7), pp.1279–91. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3137516&tool=pmcentrez&rendertype=abstract> [Accessed September 28, 2015].
- Boutet, S. et al., 2003. *Arabidopsis* HEN1. *Current Biology*, 13(10), pp.843–848. Available at: <http://www.cell.com/article/S0960982203002938/fulltext> [Accessed July 29, 2015].
- Burdisso, P. et al., 2012. Second double-stranded RNA binding domain of dicer-like ribonuclease 1: Structural and biochemical characterization. *Biochemistry*, 51(51), pp.10159–10166.
- Burdisso, P., Suarez, I.P., Bologna, N.G., Palatnik, J.F., Bersch, B. & Rasia, R.M., 2012. Second Double-Stranded RNA Binding Domain of Dicer-like Ribonuclease 1: Structural and Biochemical Characterization. *Biochemistry*, 51, pp.10159–10166.
- Cenik, E.S. et al., 2011. Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. *Molecular Cell*, 42(2), pp.172–184. Available at: <http://dx.doi.org/10.1016/j.molcel.2011.03.002>.
- Chapman, E.J. & Carrington, J.C., 2007. Specialization and evolution of endogenous small RNA pathways. *Nature reviews. Genetics*, 8(11), pp.884–96. Available at: <http://dx.doi.org/10.1038/nrg2179> [Accessed July 7, 2015].
- Chaudhury, S., Lyskov, S. & Gray, J.J., 2010. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*, 26(5), pp.689–691. Available at: <http://bioinformatics.oxfordjournals.org/content/26/5/689.abstract> [Accessed September 28, 2015].
- Chaulk, S.G. et al., 2014. MicroRNA miR-92a-1 biogenesis and mRNA targeting is modulated by a tertiary contact within the miR-17~92 microRNA cluster. *Nucleic Acids Research*, 42(8), pp.5234–5244.
- Chen, X., 2004. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science (New York, N.Y.)*, 303(5666), pp.2022–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12893888> [Accessed September 21, 2015].

- Chendrimada, T.P. et al., 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 436(7051), pp.740–4. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2944926&tool=pmcentrez&rendertype=abstract> [Accessed March 27, 2015].
- Chu, C.-Y. & Rana, T.M., 2007. Small RNAs: regulators and guardians of the genome. *Journal of cellular physiology*, 213(2), pp.412–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17674365> [Accessed September 10, 2015].
- Csermely, P., Palotai, R. & Nussinov, R., 2010. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in biochemical sciences*, 35(10), pp.539–46. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3018770&tool=pmcentrez&rendertype=abstract> [Accessed October 31, 2015].
- Cuperus, J.T. et al., 2010. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nature structural & molecular biology*, 17(8), pp.997–1003. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2916640&tool=pmcentrez&rendertype=abstract> [Accessed July 20, 2015].
- Cuperus, J.T., Fahlgren, N. & Carrington, J.C., 2011. Evolution and functional diversification of MIRNA genes. *The Plant cell*, 23(2), pp.431–442.
- Das, R. et al., 2005. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA (New York, N.Y.)*, 11(3), pp.344–54. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1262685&tool=pmcentrez&rendertype=abstract> [Accessed November 12, 2015].
- Davis, A.R. & Znosko, B.M., 2007. Thermodynamic Characterization of Single Mismatches Found in Naturally Occurring RNA. *Thermodynamic Characterization of Single Mismatches Found in Naturally Occurring RNA*. Society, pp.13425–13436.
- Delaglio, F. et al., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*, 6(3), pp.277–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8520220> [Accessed February 13, 2015].
- Dethoff, E. a. et al., 2012. Visualizing transient low-populated structures of RNA. *Nature*, 491(7426), pp.724–728. Available at: <http://dx.doi.org/10.1038/nature11498>.
- Dong, Z., Han, M.-H. & Fedoroff, N., 2008. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proceedings of the National Academy of Sciences of the United States of America*, 105(29), pp.9970–9975.
- Doyle, M., Badertscher, L. & Jaskiewicz, L., 2013. The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. *Rna*. Available at: <http://mypro.munawer.in/www.rnajournal.org/content/19/9/1238.full>.

- Du, Z. et al., 2008. Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), pp.2391–2396.
- Dunker, A.K. et al., 2001. Intrinsically disordered protein. *Journal of molecular graphics & modelling*, 19(1), pp.26–59. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11381529> [Accessed September 10, 2015].
- Dyson, H.J. et al., 2004. Unfolded Proteins and Protein Folding Studied by NMR. *Chemical Reviews*, 104(8), pp.3607–3622. Available at: <http://dx.doi.org/10.1021/cr030403s> \n<http://pubs.acs.org.proxy.ub.uni-frankfurt.de/doi/full/10.1021/cr030403s>.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–7. Available at: <http://nar.oxfordjournals.org/content/32/5/1792> [Accessed July 11, 2014].
- Ehresmann, C. et al., 1987. Probing the structure of RNAs in solution. *Nucleic acids research*, 15(22), pp.9109–28. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=306456&tool=pmcentrez&rendertype=abstract> [Accessed November 8, 2015].
- Eyal, E. et al., 2004. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *Journal of computational chemistry*, 25(5), pp.712–24. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14978714> [Accessed November 11, 2015].
- Fang, Y. & Spector, D.L., 2007. Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living Arabidopsis Plants. *Current Biology*, 17(9), pp.818–823.
- Fire, A. et al., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), pp.806–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9486653> [Accessed July 17, 2014].
- Förstemann, K. et al., 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS biology*, 3(7), p.e236. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1141267&tool=pmcentrez&rendertype=abstract> [Accessed September 10, 2015].
- Fritsch, Sambrook & Maniatis, 1989. *Molecular Cloning: A Laboratory Manual. Second Edition*, Available at: <http://www.amazon.com/Molecular-Cloning-Laboratory-Manual-EDITION/dp/B0019Y5D4O> [Accessed September 10, 2015].
- Gan, J. et al., 2008. A stepwise model for double-stranded RNA processing by ribonuclease III. *Molecular Microbiology*, 67(1), pp.143–154.
- Gan, J. et al., 2006. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, 124(2), pp.355–366.

- Gregory, B.D. et al., 2008. A link between RNA metabolism and silencing affecting Arabidopsis development. *Developmental cell*, 14(6), pp.854–66. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18486559> [Accessed September 18, 2015].
- Gregory, R.I. et al., 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014), pp.235–40. Available at: <http://dx.doi.org/10.1038/nature03120> [Accessed July 25, 2015].
- Guo, S. & Kemphues, K.J., 1995. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, 81(4), pp.611–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7758115> [Accessed August 21, 2015].
- Haase, A.D. et al., 2005. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO reports*, 6(10), pp.961–967. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1369185&tool=pmcentrez&rendertype=abstract> [Accessed September 30, 2015].
- Hamilton, A.J., 1999. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science*, 286(5441), pp.950–952. Available at: <http://www.sciencemag.org/content/286/5441/950> [Accessed September 2, 2015].
- Hammond, S.M. et al., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, 404(6775), pp.293–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10749213> [Accessed January 8, 2015].
- Han, J. et al., 2006. Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell*, 125(5), pp.887–901.
- Han, J. et al., 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes and Development*, 18(24), pp.3016–3027.
- Hanahan, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*, 166(4), pp.557–80. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6345791> [Accessed January 11, 2015].
- Hitti, E., Neunteufl, A. & Jantsch, M.F., 1998. The double-stranded RNA-binding protein Xlrpba promotes RNA strand annealing. *Nucleic Acids Research*, 26(19), pp.4382–4388.
- Huang, J.R. et al., 2013. Direct prediction of NMR residual dipolar couplings from the primary sequence of unfolded proteins. *Angewandte Chemie - International Edition*, 52(2), pp.687–690.
- lešmantavičius, V. et al., 2014. Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angewandte Chemie - International Edition*, 53(6), pp.1548–1551.
- Inada, T. & Nakamura, Y., 1995. Lethal double-stranded RNA processing activity of ribonuclease III in the absence of *suhB* protein of *Escherichia coli*. *Biochimie*, 77(4), pp.294–302. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8589060> [Accessed September 15, 2015].

- Jacobsen, S.E., Running, M.P. & Meyerowitz, E.M., 1999. Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* (Cambridge, England), 126(23), pp.5231–5243.
- Jensen, M.R. et al., 2014. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chemical reviews*, 114(13), pp.6632–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24725176>.
- Jensen, M.R. et al., 2008. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Application to the molecular recognition element of Sendai virus nucleoprotein. *Journal of the American Chemical Society*, 130(25), pp.8055–8061.
- Jensen, M.R. et al., 2009. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure*, 17(9), pp.1169–1185.
- Kabsch, W. & Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12), pp.2577–637. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6667333> [Accessed August 12, 2014].
- Kapust, R.B. et al., 2001. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein engineering*, 14(12), pp.993–1000. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11809930> [Accessed June 24, 2015].
- Katiyar-Agarwal, S. et al., 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), pp.18002–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693862&tool=pmcentrez&rendertype=abstract> [Accessed September 10, 2015].
- Kibbe, W.A., 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic acids research*, 35(Web Server issue), pp.W43–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1933198&tool=pmcentrez&rendertype=abstract> [Accessed February 19, 2015].
- Kim, V.N., 2005. Small RNAs: classification, biogenesis, and function. *Molecules and cells*, 19(1), pp.1–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15750334> [Accessed September 28, 2015].
- Kim, V.N., Han, J. & Siomi, M.C., 2009. Biogenesis of small RNAs in animals. *Nature reviews. Molecular cell biology*, 10(2), pp.126–39. Available at: <http://dx.doi.org/10.1038/nrm2632> [Accessed July 9, 2014].
- Kleckner, I.R. & Foster, M.P., 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta*, 1814(8), pp.942–68. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3061256&tool=pmcentrez&rendertype=abstract> [Accessed August 27, 2015].
- Kok, K.H. et al., 2007. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *The Journal of biological*

- chemistry, 282(24), pp.17649–57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17452327> [Accessed September 16, 2015].
- Kowalinski, E. et al., 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, 147(2), pp.423–35. Available at: <http://www.cell.com/article/S0092867411011500/fulltext> [Accessed September 19, 2015].
- Kriwacki, R.W. et al., 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21), pp.11504–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=38087&tool=pmcentrez&rendertype=abstract> [Accessed October 31, 2015].
- Kurihara, Y., Takashi, Y. & Watanabe, Y., 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA (New York, N.Y.)*, 12(2), pp.206–212.
- Kurihara, Y. & Watanabe, Y., 2004. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), pp.12753–12758.
- Laederach, A. et al., 2009. *NIH Public Access*, 3(9), pp.1395–1401.
- Laederach, A. et al., 2008. Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nature protocols*, 3(9), pp.1395–401. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2652576&tool=pmcentrez&rendertype=abstract> [Accessed September 12, 2015].
- Lamontagne, B. & Elela, S.A., 2004. Evaluation of the RNA Determinants for Bacterial and Yeast RNase III Binding and Cleavage. *Journal of Biological Chemistry*, 279(3), pp.2231–2241.
- Lau, P.-W. et al., 2012. The molecular architecture of human Dicer. *Nature Structural & Molecular Biology*, 19(4), pp.436–440. Available at: <http://dx.doi.org/10.1038/nsmb.2268>.
- Laubinger, S. et al., 2008. Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 105(25), pp.8795–8800.
- Lee, Y. et al., 2006. The role of PACT in the RNA silencing pathway. *The EMBO journal*, 25(3), pp.522–32. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1383527&tool=pmcentrez&rendertype=abstract> [Accessed August 27, 2015].
- Lemieux, S. & Major, F., 2002. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic acids research*, 30(19), pp.4250–4263.
- Lescop, E., Schanda, P. & Brutscher, B., 2007. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 187(1), pp.163–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17468025> [Accessed July 29, 2015].

- Leulliot, N. et al., 2004. A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *The EMBO journal*, 23(13), pp.2468–77. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=449770&tool=pmcentrez&rendertype=abstract> [Accessed November 3, 2015].
- Lightfoot, H.L. et al., 2011. A LIN28-dependent structural change in pre-let-7g directly inhibits dicer processing. *Biochemistry*, 50(35), pp.7514–7521.
- Liu, C., Axtell, M.J. & Fedoroff, N. V., 2012. The Helicase and RNaseIIIa Domains of Arabidopsis Dicer-Like1 Modulate Catalytic Parameters during MicroRNA Biogenesis. *Plant Physiology*, 159(2), pp.748–758.
- Liu, Q. et al., 2013. Complementation of Hyponastic Leaves1 by Double-Strand RNA-Binding Domains of Dicer-Like1 in Nuclear Dicing Bodies. *Plant Physiology*, 163(1), pp.108–117. Available at: <http://www.plantphysiol.org/cgi/doi/10.1104/pp.113.219071>.
- Liu, Q., Feng, Y. & Zhu, Z., 2009. Dicer-like (DCL) proteins in plants. *Functional and Integrative Genomics*, 9(3), pp.277–286.
- Llave, C. et al., 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science (New York, N.Y.)*, 297(5589), pp.2053–6. Available at: <http://www.sciencemag.org/content/297/5589/2053.long> [Accessed August 6, 2015].
- Lobbes, D. et al., 2006. SERRATE: a new player on the plant microRNA scene. *EMBO reports*, 7(10), pp.1052–1058.
- Louhivuori, M. et al., 2003. On the origin of residual dipolar couplings from denatured proteins. *Journal of the American Chemical Society*, 125(50), pp.15647–15650. Available at: <http://pubs.acs.org/doi/abs/10.1021/ja035427v> \npapers2://publication/doi/10.1021/ja035427v.
- Lowman, H.B. & Draper, D.E., 1986. On the recognition of helical RNA by cobra venom V1 nuclease. *Journal of Biological Chemistry*, 261(12), pp.5396–5403.
- Ma, E. et al., 2012. Coordinated activities of human dicer domains in regulatory RNA processing. *Journal of Molecular Biology*, 422(4), pp.466–476. Available at: <http://dx.doi.org/10.1016/j.jmb.2012.06.009>.
- Macrae, I.J. et al., 2006. Structural basis for double-stranded RNA processing by Dicer. *Science (New York, N.Y.)*, 311(5758), pp.195–198.
- MacRae, I.J. & Doudna, J. a., 2007. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current Opinion in Structural Biology*, 17(1), pp.138–145.
- Manavella, P. a. et al., 2012. Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell*, 151(4), pp.859–870.
- Margis, R. et al., 2006. The evolution and diversification of Dicers in plants. *FEBS Letters*, 580(10), pp.2442–2450.

- Markley, J.L. et al., 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *Journal of biomolecular NMR*, 12(1), pp.1–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9729785> [Accessed September 13, 2015].
- Marsh, J. a et al., 2006. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein science : a publication of the Protein Society*, 15(12), pp.2795–2804.
- Marsh, J.A. & Forman-kay, J.D., 2010. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophysj*, 98(10), pp.2383–2390. Available at: <http://dx.doi.org/10.1016/j.bpj.2010.02.006>.
- Maslah, G., Barraud, P. & Allain, F.H.T., 2013. RNA recognition by double-stranded RNA binding domains: A matter of shape and sequence. *Cellular and Molecular Life Sciences*, 70(11), pp.1875–1895.
- Mateos, J.L. et al., 2010. Identification of MicroRNA Processing Determinants by Random Mutagenesis of Arabidopsis MIR172a Precursor. *Current Biology*, 20(1), pp.49–54.
- Mathews, D.H. et al., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*, 288(5), pp.911–940.
- Mello, C.C. & Conte, D., 2004. Revealing the world of RNA interference. *Nature*, 431(7006), pp.338–42. Available at: <http://dx.doi.org/10.1038/nature02872> [Accessed September 25, 2015].
- Mohana-Borges, R. et al., 2004. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *Journal of molecular biology*, 340(5), pp.1131–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15236972> [Accessed November 2, 2015].
- Montgomery, T.A. et al., 2008. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, 133(1), pp.128–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18342362> [Accessed September 26, 2015].
- Napoli, C., Lemieux, C. & Jorgensen, R., 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, 2(4), pp.279–289. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=159885&tool=pmcentrez&rendertype=abstract> [Accessed September 16, 2015].
- Narayanan, R. et al., 2008. Kinetics of Folding and Binding of an Intrinsically Disordered Protein : The Inhibitor of Yeast Aspartic Proteinase YPrA Kinetics of Folding and Binding of an Intrinsically Disordered Protein : The Inhibitor of Yeast Aspartic Proteinase YPrA. *Design*, 97(12), pp.11477–11485.
- Nicholson, A.W., 2014. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdisciplinary Reviews: RNA*, 5(1), pp.31–48.



- Nodine, M.D. & Bartel, D.P., 2010. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes & development*, 24(23), pp.2678–92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21123653> [Accessed September 22, 2015].
- Oldfield, C.J. et al., 2008. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*, 9(Suppl 1), p.S1. Available at: <http://www.biomedcentral.com/1471-2164/9/S1/S1>.
- Ozenne, V. et al., 2012. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *Journal of the American Chemical Society*, 134(36), pp.15138–48. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22901047>.
- Park, M.Y. et al., 2005. Nuclear processing and export of microRNAs in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10), pp.3691–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=553294&tool=pmcentrez&rendertype=abstract> [Accessed September 13, 2015].
- Park, W. et al., 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Current biology : CB*, 12(17), pp.1484–95. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12225663> [Accessed September 26, 2015].
- Peragine, A. et al., 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes & development*, 18(19), pp.2368–79. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=522987&tool=pmcentrez&rendertype=abstract> [Accessed September 26, 2015].
- Pertzev, A. V. & Nicholson, A.W., 2006. Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of Escherichia coli ribonuclease III. *Nucleic Acids Research*, 34(13), pp.3708–3721.
- Price, M.N., Dehal, P.S. & Arkin, A.P., 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one*, 5(3), p.e9490. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009490> [Accessed July 10, 2014].
- Qin, H. et al., 2010. Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA (New York, N.Y.)*, 16(3), pp.474–481.
- Quick-Cleveland, J. et al., 2014. The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. *Cell Reports*, 7(6), pp.1994–2005. Available at: <http://dx.doi.org/10.1016/j.celrep.2014.05.013>.
- Rasia, R.M. et al., 2010. Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry*, 49(38), pp.8237–8239.

- Ren, G. et al., 2012. Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(31), pp.12817–21. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3412041&tool=pmcentrez&rendertype=abstract> [Accessed September 30, 2015].
- Romano, N. & Macino, G., 1992. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Molecular microbiology*, 6(22), pp.3343–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1484489> [Accessed August 18, 2015].
- Rückert, M. & Otting, G., 2000. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *Journal of the American Chemical Society*, 122(32), pp.7793–7797. Available at: <http://dx.doi.org/10.1021/ja001068h> [Accessed July 20, 2015].
- Saito, K. et al., 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS biology*, 3(7), p.e235. Available at: <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030235> [Accessed September 30, 2015].
- Schanda, P. & Brutscher, B., 2005. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22), pp.8014–5. Available at: <http://dx.doi.org/10.1021/ja051306e> [Accessed September 13, 2015].
- Schauer, S.E. et al., 2002. DICER-LIKE1: Blind men and elephants in *Arabidopsis* development. *Trends in Plant Science*, 7(11), pp.487–491.
- Shen, Y. et al., 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12), pp.4685–4690.
- Shen, Y. & Bax, A., 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR*, 48(1), pp.13–22. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2935510&tool=pmcentrez&rendertype=abstract> [Accessed September 27, 2015].
- Shi, Z. et al., 2011. Characterization of *Aquifex aeolicus* ribonuclease III and the reactivity epitopes of its pre-ribosomal RNA substrates. *Nucleic Acids Research*, 39(7), pp.2756–2768.
- Shoemaker, B.A., Portman, J.J. & Wolynes, P.G., 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp.8868–73. Available at: <http://www.pnas.org/content/97/16/8868.full> [Accessed October 31, 2015].
- Song, L. et al., 2007. *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp.5437–5442.

- Song, L., Axtell, M.J. & Fedoroff, N. V., 2010a. RNA Secondary Structural Determinants of miRNA Precursor Processing in Arabidopsis. *Current Biology*, 20(1), pp.37–41. Available at: <http://dx.doi.org/10.1016/j.cub.2009.10.076>.
- Song, L., Axtell, M.J. & Fedoroff, N. V., 2010b. RNA Secondary Structural Determinants of miRNA Precursor Processing in Arabidopsis. *Current Biology*, 20(1), pp.37–41.
- Stefl, R. et al., 2010. The Solution Structure of the ADAR2 dsRBM-RNA Complex Reveals a Sequence-Specific Readout of the Minor Groove. *Cell*, 143(2), pp.225–237.
- Sugase, K., Dyson, H.J. & Wright, P.E., 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147), pp.1021–1025.
- Tavares, T.J., Beribisky, A. V & Johnson, P.E., 2009. Structure of the cytosine-cytosine mismatch in the thymidylate synthase mRNA binding site and analysis of its interaction with the aminoglycoside paromomycin. *RNA (New York, N.Y.)*, 15(5), pp.911–922.
- Taylor, D.W. et al., 2013. Substrate-specific structural rearrangements of human Dicer. *Nature structural & molecular biology*, 20(6), pp.662–70. Available at: <http://dx.doi.org/10.1038/nsmb.2564> [Accessed September 13, 2015].
- Tian, B. et al., 2004. The double-stranded-RNA-binding motif: interference and much more. *Nature reviews. Molecular cell biology*, 5(12), pp.1013–1023.
- Tian, Y. et al., 2014. A Phosphate-Binding Pocket within the Platform-PAZ-Connector Helix Cassette of Human Dicer. *Molecular Cell*, 53(4), pp.606–616. Available at: <http://dx.doi.org/10.1016/j.molcel.2014.01.003>.
- Tolman, J.R. & Ruan, K., 2006. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chemical reviews*, 106(5), pp.1720–36. Available at: <http://dx.doi.org/10.1021/cr040429z> [Accessed November 12, 2015].
- Touw, W.G. et al., 2015. A series of PDB-related databanks for everyday needs. *Nucleic acids research*, 43(Database issue), pp.D364–8. Available at: <http://nar.oxfordjournals.org/content/43/D1/D364> [Accessed August 12, 2015].
- Tsutsumi, A. et al., 2011. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nature Structural & Molecular Biology*, 18(10), pp.1153–1158.
- Turjanski, A.G. et al., 2008. Binding-induced folding of a natively unstructured transcription factor. *PLoS computational biology*, 4(4), p.e1000060. Available at: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000060> [Accessed November 3, 2015].
- Ucci, J.W. et al., 2007. Mechanism of interaction of the double-stranded RNA (dsRNA) binding domain of protein kinase R with short dsRNA sequences. *Biochemistry*, 46(1), pp.55–65.
- Uversky, V.N., 2013. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*, 22(6), pp.693–724.

- Uversky, V.N., 2015. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS letters*, 589(19), pp.2498–2506. Available at: <http://www.sciencedirect.com/science/article/pii/S0014579315004603> [Accessed June 18, 2015].
- Uversky, V.N., Oldfield, C.J. & Dunker, a. K., 2005. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*, 18(5), pp.343–384.
- Vazquez, F., 2006. *Arabidopsis* endogenous small RNAs: highways and byways. *Trends in Plant Science*, 11(9), pp.460–468.
- Vazquez, F. et al., 2004. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Current Biology*, 14(4), pp.346–351.
- Vranken, W.F. et al., 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59(4), pp.687–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15815974> [Accessed October 20, 2014].
- Vukovic, L. et al., 2014. Substrate recognition and specificity of double-stranded RNA binding proteins. , (May).
- Wang, Y. & Jardetzky, O., 2002. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein science : a publication of the Protein Society*, 11(4), pp.852–61. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2373532&tool=pmcentrez&rendertype=abstract> [Accessed July 21, 2015].
- Weinberg, D.E. et al., 2011. The inside-out mechanism of dicers from budding yeasts. *Cell*, 146(2), pp.262–276. Available at: <http://dx.doi.org/10.1016/j.cell.2011.06.021>.
- Welker, N.C. et al., 2011. Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode. *Molecular Cell*, 41(5), pp.589–599. Available at: <http://dx.doi.org/10.1016/j.molcel.2011.02.005>.
- Werner, S. et al., 2010. Structure Determinants for Accurate Processing of miR172a in *Arabidopsis thaliana*. *Current Biology*, 20(1), pp.42–48. Available at: <http://dx.doi.org/10.1016/j.cub.2009.10.073>.
- Wilkins, D.K. et al., 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, 38(50), pp.16424–16431.
- Willmann, M.R. et al., 2011. MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. *Plant physiology*, 155(4), pp.1871–1884.
- Wishart, D.S. et al., 1995. <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, 5(1), pp.67–81. Available at: <http://link.springer.com/10.1007/BF00227471> [Accessed May 27, 2015].

- Wostenberg, C. et al., 2012. The Role of Human Dicer-dsRBD in Processing Small Regulatory RNAs. , 7(12).
- Wostenberg, C., Quarles, K.A. & Showalter, S.A., 2010. Dynamic Origins of Differential RNA Binding Function in Two dsRBDs from the miRNA "Microprocessor" Complex. *Biochemistry*, 49(50), pp.10728–10736. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3565223&tool=pmcentrez&rendertype=abstract> [Accessed October 1, 2015].
- Wright, P.E. & Dyson, H.J., 2009. Linking folding and binding. *Current opinion in structural biology*, 19(1), pp.31–8. Available at: <http://www.sciencedirect.com/science/article/pii/S0959440X08001796> [Accessed December 10, 2014].
- Wu, X. et al., 2013. A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. *Cell research*, 23(5), pp.645–57. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641593&tool=pmcentrez&rendertype=abstract>.
- Xie, Z. et al., 2005. Expression of Arabidopsis MIRNA genes. *Plant physiology*, 138(4), pp.2145–54. Available at: <http://www.plantphysiol.org/content/138/4/2145> [Accessed September 28, 2015].
- Xie, Z., Khanna, K. & Ruan, S., 2010. Expression of microRNAs and its regulation in plants. *Seminars in Cell and Developmental Biology*, 21(8), pp.790–797. Available at: <http://dx.doi.org/10.1016/j.semcdb.2010.03.012>.
- Xie, Z. & Qi, X., 2008. Diverse small RNA-directed silencing pathways in plants. *Biochimica et biophysica acta*, 1779(11), pp.720–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18358247> [Accessed September 28, 2015].
- Yang, L. et al., 2006. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant Journal*, 47(6), pp.841–850.
- Yu, B. et al., 2008. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(29), pp.10073–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2481372&tool=pmcentrez&rendertype=abstract> [Accessed September 30, 2015].
- Zamore, P.D. et al., 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1), pp.25–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10778853> [Accessed July 10, 2015].
- Zhan, X. et al., 2012. Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), pp.18198–203. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3497810&tool=pmcentrez&rendertype=abstract> [Accessed September 30, 2015].

- Zhang, H. et al., 2004. Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1), pp.57–68.
- Zheng, X. & Bevilacqua, P.C., 2000. Straightening of bulged RNA by the double-stranded RNA-binding domain from the protein kinase PKR. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26), pp.14162–14167.
- Zhu, H. et al., 2013. Bidirectional processing of pri-miRNAs with branched terminal loops by *Arabidopsis* Dicer-like1. *Nature structural & molecular biology*, 20(9), pp.1106–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23934148>.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), pp.3406–3415.
- Zweckstetter, M. & Bax, A., 2000. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *Journal of the American Chemical Society*, 122(15), pp.3791–3792. Available at: <http://dx.doi.org/10.1021/ja0000908> [Accessed October 28, 2015].